Theta phase synchronization is the glue that binds human associative memory

Clouter, Andrew; Shapiro, Kimron; Hanslmayr, Simon

[Link to publication on Research at Birmingham portal](#)

Title: Theta phase synchronization is the glue that binds human associative memory

Authors: Andrew Clouter[1], Kimron L. Shapiro[1], Simon Hanslmayr[1,2,*]

Affiliations:

[1] School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

5   [2] Lead Contact

* Correspondence: s.hanslmayr@bham.ac.uk

Summary

Episodic memories are information-rich, often multisensory events that rely on binding different

10   elements [1]. The elements that will constitute a memory episode are processed in specialized but

distinct brain modules. The binding of these elements is likely mediated by fast-acting long-term

potentiation (LTP), which relies on the precise timing of neural activity [2]. Theta oscillations in the

hippocampus orchestrate such timing as demonstrated by animal studies in vitro [3, 4] and in vivo

[5, 6] suggesting a causal role of theta activity for the formation of complex memory episodes, but

15   direct evidence from humans is missing. Here we show that human episodic memory formation

depends on phase synchrony between different sensory cortices at the theta frequency. By

modulating the luminance of visual stimuli and the amplitude of auditory stimuli, we directly

manipulated the degree of phase synchrony between visual and auditory cortices. Memory for

sound-movie associations was significantly better when the stimuli were presented in-phase

20   compared to out-of-phase. This effect was specific to theta (4 Hz), and did not occur in slower (1.7

Hz) or faster (10.5 Hz) frequencies. These findings provide the first direct evidence that episodic

memory formation in humans relies on a theta-specific synchronization mechanism.

Results

In order for the sensory elements that originate from different brain regions to be bound into an episodic memory, the different information processing pathways must converge at a site where this binding takes place (such as the hippocampus [7-9]). Theta oscillations, which provide time windows for fast-acting LTP and LTD [3-6], likely act as the "gluing mechanism" for human memories [10, 11]. Corroborating evidence from human studies suggests that theta oscillatory (~4 Hz) activity in the medial temporal lobe (MTL) correlates specifically with binding in episodic memory [12]. To provide direct evidence that theta synchronisation is the "glue" that enables episodic memories to be created from sensory input, we devised a novel multisensory memory paradigm to achieve three goals: (1) To demonstrate that the degree of synchrony (i.e., the degree of phase offset) between visual and auditory stimuli during memory formation (i.e., encoding) has an impact on later recall; (2) To demonstrate that activity in auditory and visual brain areas is responsive to oscillatory synchronisation of the sensory input; and (3) To establish that this effect is specific to the theta frequency band.

Episodic memory associations are modulated by theta phase synchrony

We employed a novel multisensory memory paradigm to control the degree of synchrony between visual and auditory stimuli (Figure 1). Specifically, we presented short movie clips together with unrelated music clips (see STAR Methods; see Audio S1 and Movie S1 for examples). The movie and sound clips were modulated separately by applying a sine wave, such that the luminance of the movie and the amplitude of the sound fluctuated from 0% to 100% at a chosen frequency, and with a particular phase angle, which allowed us to achieve synchrony or asynchrony. The sine waves modulating the movie luminance and the sound amplitude were either in-phase (i.e., 0° phase offset) or out-of-phase by a 90°, 180°, or 270° offset. Given that neural processing time differs between the senses (i.e., retinal phototransduction is a relatively slow process compared to auditory transduction, taking approximately 50 vs. 10 ms, respectively [13-

19]), the stimuli were modulated such that the onset of the movies led the onset of the sound clips by approximately 40 ms (i.e. ~58 degrees at 4 Hz). The phase offsets reported above thus refer to the hypothesised phase differences in the brain. The frequency of interest here was 4 Hz—the human equivalent of the hippocampal theta recorded in rodents [20]. It has been shown that presenting each stimulus in such a fashion leads to strong rhythmic modulations of the targeted sensory areas that are phase locked to the stimulation (i.e., entrained [21, 22]).

To achieve our first goal, i.e., to demonstrate that the degree of synchrony between visual and auditory stimuli during memory formation (i.e., encoding) modulates recall, we tested 24 participants using a memory task that requires cross-domain associative binding and thus strongly depends on hippocampal function. The task required remembering associations between pairs of short movies and sound clips (3 s duration each). During the encoding phase participants rated how well the sound suited the movie after each presentation to draw attention to both modalities. A distraction task following each block (counting backwards for 30 s) served to prevent rehearsal. Participants then completed an associative recall test; on each trial one of the sounds from the encoding phase was played (Figure S1A) and participants were required to select from a set of four movies (all of which were seen in the same encoding phase) the movie that they recalled as playing when they heard that same sound during the encoding phase (Figure S1B). Six blocks of such encoding-retrieval cycles (16 trials per cycle) were conducted, with each block containing randomly ordered trials at each phase offset (see STAR Methods).

The results of the first experiment for each of the four phase offset conditions are shown in Figure 2A. The effect of phase offset on memory recall was significant as evidenced by a repeated measures ANOVA: $F(3,69) = 6.74$, $P < 0.01$. Pairwise t-tests confirmed that memory performance was better in the 0° phase offset condition than in the 90°, 180°, and 270° phase offset conditions ($T(1,23) = 3.456$, $P = 0.01$, $T(1,23) = 3.915$, $P < 0.01$, and $T(1,23) = 4.281$, $P < 0.01$, respectively). Additional pairwise t-tests confirmed that memory performance in the 90°, 180°, and 270° phase offset conditions were not statistically different from each other (all $P > 0.05$). A control experiment

where the flickering condition was compared to a non-flickering baseline revealed that this effect was not driven by perceptual factors (see STAR Methods). We also measured the ability of participants to discriminate synchronous from asynchronous stimulus presentation: Sensitivity

80 analysis (d-prime) revealed that judgement of synchrony was no better for synchronous or asynchronous conditions, i.e., participants could not discriminate between synchronous and asynchronous stimuli, suggesting that perceptual judgements of synchrony were not responsible for performance on the episodic memory task (Figure 2B). Together, the results of this first experiment support our first goal to reveal the striking finding that synchronised sensory stimuli increase

85 human associative memory.

Activity in sensory brain areas is responsive to oscillatory synchronisation of the sensory input

Our second goal was to demonstrate that modulation of the auditory and visual stimuli used

90 in our experiments modulates in a corresponding way the respective brain processing areas. The finding that presenting stimuli in a rhythmic fashion modulates the respective sensory area, which becomes phase locked to the stimulation, is not new [21, 22]. Nevertheless, for our study it was important to ensure that we were successful in inducing oscillatory activity in the visual and auditory cortices at pre-specified phase angles. To achieve this, we recorded electrical activity in the

95 brain using electroencephalography (EEG) during the presentation of sine wave modulated sound-movie clips. EEG data were collected for nine participants, each of whom completed a task that required rating how well a series of short sound clips suited the content of short movie clips. Twelve blocks of the rating task (16 trials per block) were conducted (the "multimodal conditions"). Following the 12 rating task blocks, participants also completed two "unimodal" stimulation blocks

100 that involved rating how pleasant each short sound clip or each movie clip was on its own (each block contained 50 trials, containing previously presented movie and sound clips). Source localisation analysis of the EEG data from the unimodal conditions enabled identification of dipole

sources in the auditory and visual cortices; the EEG data at those sources were then reconstructed in the multimodal conditions (see STAR Methods).

105   EEG waveforms at auditory and visual sources reveal phase entrainment corresponding to their respective phase offsets, as shown in Figure 3 and Figure S2. To calculate the probability that the phase difference between the signals at the auditory and visual sources was uniformly distributed (i.e., no phase synchrony), Rayleigh's uniformity test and the V test [23] were conducted on the grand average EEG signals between 0.75 and 2.75 seconds (thus excluding onset and offset

110 responses; 1026 samples in each phase offset condition). In each phase offset condition, Rayleigh's test rejected the hypothesis of uniformly distributed phase differences, and the V test rejected uniformly distributed phase differences in favour of the relevant alternative hypothesis (i.e., that the mean phase difference was 0°, 90°, 180°, or 270°). Table S1 shows the results of the circular analyses. These results enable us to conclude that auditory and visual brain regions were entrained

115 by their respective sensory stimuli, revealing success in achieving our second goal.

Theta phase synchrony specifically modulates associative episodic memory formation

      To achieve our third goal, i.e., to establish that this effect is specific to modulation in the theta frequency band, revealed by the literature to be fundamental to associative memory, we sought

120 to examine whether the improved memory performance for the synchronous (i.e., 0° phase offset) compared to the asynchronous (90°, 180°, 270° phase offset) condition was specific to 4 Hz. We repeated the first experiment with 24 new participants, using the same movie and sound stimuli; however in this experiment the movies and sounds were modulated either with a 4 Hz (~theta), 1.7 Hz (~delta) or 10.5 Hz (~alpha) sine wave (see STAR Methods). These frequencies were chosen

125 such that they would not share harmonics with 4 Hz, a crucial requirement when testing for frequency specificity [24]. On half of the trials for each frequency, we modulated the movie luminance and sound amplitude to be in-phase (the synchronous condition), whereas on the other

half of trials the modulation was offset by 90°, 180°, or 270° (collectively the asynchronous condition).

130   Memory performance and differences between the synchronous and asynchronous conditions for each of the three frequencies are shown in Figures 2C and 2D. In the theta condition memory performance was better for the synchronous compared to the asynchronous condition: $T(1,23) = 3.34$, $P < 0.01$, which replicates the results from the previous experiment (note that due to an unequal number of trials in each condition, the synchronous condition was compared directly to

135   the combined asynchronous conditions, rather than to the three asynchronous phase offsets). No significant differences between the synchronous and asynchronous conditions emerged in the other two frequency conditions (delta: $T(1,23) = 0.07$, $P = 0.95$; alpha: $T(1,23) = 0.17$, $P = 0.87$). Additionally, a direct comparison of the synchronous-asynchronous differences between theta and the other two frequencies showed that the synchrony effect at theta was significantly greater than

140   the combined effect in the two control frequencies: $T(1,23) = 2.46$, $P = 0.01$ (one-sided). Moreover, memory performance for synchronously presented stimuli was better in the theta condition than in the two control frequencies: $T(1,23) = 1.93$, $P = 0.03$ (one-sided). As before, control analyses confirmed that the effect at the theta frequency was not driven by spurious perceptual effects (see STAR Methods and Figure 2B). Importantly, it is clear that synchronous stimulus presentation at

145   theta improved associative memory, rather than asynchronous presentation impairing memory. The stimuli were equally distracting (flickering) within each frequency condition, yet synchronous presentation at theta yielded better memory performance than synchronous presentation at other frequencies. We thus achieved our second goal by showing that episodic memory associations are specifically modulated by theta phase synchrony, but not by either slower or faster frequencies.

150   Accordingly we term this effect the Theta Induced Memory Effect (TIME).

Discussion

Our experiments reveal the novel and important finding that human episodic memory relies on precision coordinated timing of real world stimuli. We found that stimuli presented

155 synchronously (at the same time and phase) and modulated at a particular frequency facilitates episodic memory. Modulating our stimuli at 4 Hz or human theta, a rhythm which has been associated with a region of the human brain known to be involved in long-term memory [13], was implemented to ensure that the respective brain regions would oscillate correspondingly, although it is known that non-modulated stimuli induce internal oscillations [10]. Intriguingly, subtle

160 differences in phase synchrony between auditory and visual stimuli (i.e., +/- 125 milliseconds), which participants were unable to consciously perceive (Figure 2B), strongly influenced the degree to which multisensory memories were formed. This effect is consistent and replicable as demonstrated in the present series of experiments. Importantly, the effect of phase synchrony on associative multisensory memory formation was observed only at theta, but not at a slower (delta)

165 or a faster (alpha) frequency. Moreover, memory improved not only relative to asynchronously presented stimuli, but also relative to non-flickering stimuli (see the STAR Methods). Our results therefore provide direct evidence for a causal role of theta phase synchronization for the formation of human memories. In addition to elucidating a mechanism for the formation of associations in human memory, the results of our paradigm suggest possible interventions for improving memory

170 in at-risk or patient populations.

The present series of experiments modulated the auditory and visual stimuli using flicker. Given that perceptual binding is better when flickering stimuli are presented synchronously versus asynchronously [25], our pattern of results in principle could be explained by this account, rather than solely by associative memory. In a control experiment (see STAR Methods) we compared

175 memory for stimuli modulated with a 4 Hz theta sine wave to unmodulated (i.e., non-flickering) stimuli. Results show better memory performance for the 0-degrees phase offset stimuli than the non-flickering stimuli (Figure S4). Given that non-flickering stimuli present a best case scenario for

binding at the perceptual level [26], our results suggest that the theta-induced memory effect (TIME) is indeed a memory effect, and not an effect resulting from perceptual binding.

180   It has been argued that synaptic modification, the neural basis of learning and memory, relies on finely-tuned timing of the involved neural assemblies in the theta frequency range [4, 11, 27, 28]. The importance of the relative timing of events in learning and memory has been known since the pioneering work of Pavlov [29], who showed that the strength of conditioning was a function of the time between the conditioned and to-be-conditioned stimuli. Given nonlinear

185 dendritic processing, coincident inputs from sensory regions to a target neural population (e.g., the hippocampus) will generate a significantly greater response, and thus greater LTP, than non-coincident inputs [30, 31]. We did not directly measure neural activity in the hippocampus in the present series of studies, and therefore it remains an open question as to whether the present paradigm indeed entrains neural assemblies in this brain region. Nevertheless, our results are

190 consistent with the idea that theta phase represents discrete time windows for synaptic plasticity. An alternate explanation is that the modulated stimuli used in our experiments created an entrained state conducive to encoding associations. However, such an explanation would predict that memory performance would be better in the 90-degree and 270-degree phase offset conditions than the 180-degree phase offset condition, since a summation of the entraining waveforms from the two

195 modalities would cancel each other out in the 180-degree condition, but would not do so in the 90- and 270-degree conditions. Given that we did not see any differences in memory performance between the three asynchronous phase offset conditions, we believe that our observed effect is a result of the coincident timing of the inputs in the synchronous condition. While the question of which specific regions of the brain are primarily responsible for the theta-induced memory effect

200 remains open, our results bridge an important gap by providing a link between oscillatory phase synchrony and episodic memory in humans.

215

220

225

References:

1. Tulving, E. (2002). Episodic memory: from mind to brain. Annu Rev Psychol. *53*(1), 1-25.

2. Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science. *275*(5297), 213-215.

3. Pavlides, C., Greenstein, Y. J., Grudman, M., & Winson, J. (1988). Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of θ-rhythm. Brain Res. *439*(1), 383-387.

4. Huerta, P. T., & Lisman, J. E. (1995). Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro. Neuron, *15*(5), 1053-1063.

5. Hyman, J. M., Wyble, B. P., Goyal, V., Rossi, C. A., & Hasselmo, M. E. (2003). Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to the peak of theta and long-term depression when delivered to the trough. J Neurosci. *23*(37), 11725-11731.

6. Wittenberg, G. M., & Wang, S. S. H. (2006). Malleability of spike-timing-dependent plasticity at the CA3–CA1 synapse. J Neurosci. *26*(24), 6610-6617.

7. Marr, D., Willshaw, D., & McNaughton, B. (1991). Simple memory: a theory for archicortex. In From the Retina to the Neocortex. (Boston: Birkhäuser), pp. 59-128.

250   8. Moscovitch, M. (2008). The hippocampus as a "stupid," domain-specific module: Implications for theories of recent and remote memory, and of imagination. Can J Exp Psychol. *62*(1), 62.

9. Staresina, B. P., & Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. Neuron, *63*(2), 267-276.

255

10. Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. Science, *304*(5679), 1926-1929.

11. Hanslmayr, S., Staresina, B. P., & Bowman, H. (2016). Oscillations and episodic memory:
260   Addressing the synchronization/desynchronization conundrum. Trends Neurosci. *39*(1), 16-25.

12. Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. Curr Biol. *23*(12), 1101-1106.

265 13. Bolz, J., Rosner, G., & Wässle, H. (1982). Response latency of brisk- sustained (X) and brisk-transient (Y) cells in the cat retina. J Physio. *328*(1), 171-190.

14. Lamb, T. D., & Pugh, E. N. (1992). A quantitative account of the activation steps involved in phototransduction in amphibian photoreceptors. J Physio. *449*(1), 719-758.

270

15. Lennie, P. (1981). The physiological basis of variations in visual latency. Vision Res. *21*(6), 815-824.

16. Rodieck, R. W., & Rodieck, R. W. (1998). The first steps in seeing (Vol. 1). (Sunderland, MA: Sinauer Associates).

17. Schnapf, J. L., Kraft, T. W., & Baylor, D. A. (1987). Spectral sensitivity of human cone photoreceptors. Nature, *325*(6103), 439-441.

18. Corey, D. P., & Hudspeth, A. J. (1979). Response latency of vertebrate hair cells. Biophys J. *26*(3), 499-506.

19. King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. Exp Brain Res. *60*(3), 492-500.

20. Jacobs, J. (2014). Hippocampal theta oscillations are slower in humans than in rodents: implications for models of spatial navigation and memory. Phil. Trans. R. Soc. B, *369*(1635), 20130304.

21. Pantev, C., Roberts, L. E., Elbert, T., Roβ, B., & Wienbruch, C. (1996). Tonotopic organization of the sources of human auditory steady-state responses. Hearing Res. *101*(1), 62-74.

22. Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. Rev. Pap. Int. J. Audiol. *42*(4), 177-219.

23. Zar. J. H. (1979). Biostatistical Analysis (Englewood Cliffs, NJ: Prentice-Hall).

24. Pletzer, B., Kerschbaum, H., & Klimesch, W. (2010). When frequencies never synchronize: the golden mean and the resting EEG. Brain Res. *1335*, 91-102.

25. Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. Curr. Biol. *22*(1), 46-49.

26. VanRullen, R., Zoefel, B., & Ilhan, B. (2014). On the cyclic nature of perception in vision versus audition. Phil Trans R Soc B. *369*(1641), 1-14.

27. Hasselmo, M. E., Bodelón, C., & Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. Neural Comput. *14*(4), 793-817.

28. Hasselmo, M. E. (2005). What is the function of hippocampal theta rhythm?—Linking behavioral data to phasic properties of field potential and unit recording data. Hippocampus, *15*(7), 936-949.

29. Pavlov, I. P. (1927). Conditioned Reflexes (London: Oxford University Press).

30. Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. Neuron, *68*(3), 362-385.

31. Fries, P. (2015). Rhythms for cognition: communication through coherence. Neuron, *88*(1), 220-235.

32. Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. Spatial Vision, *10*, 433-436.

325

33. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, *10*(4), 437-442.

34. Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's

330   new in Psychtoolbox-3. Perception, *36*(14), 1.

35. Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci., 1.

335

36. Murzin, V., Fuchs, A., & Kelso, J. S. (2013). Detection of correlated sources in EEG using combination of beamforming and surface Laplacian methods. J Neurosci Methods. *218*(1), 96-102.

340   37. Pinheiro, J., & Bates, D. (2006). Mixed-effects models in S and S-PLUS (New York: Springer).

38. Akaike, H. (1974). A new look at the statistical model identification. IEEE T Automat Contr. *19*(6), 716-723.

345   39. Royall, R. (1997). Statistical evidence: a likelihood paradigm (Vol. 71). (London: Chapman & Hall, London).

Figure Legends

Figure 1. Depiction of the experimental paradigm.

Sinusoidally flickering visual (red) and auditory (blue) stimuli are presented either in synchrony, or at 90°, 180°, or 270° phase offset.

Figure 2. Performance on the associative memory task.

(A) Accuracy (%) of selecting the correct movies that were presented with particular sounds, when the movies and sounds were presented at each phase offset in experiment 1.

(B) Sensitivity index (d`) on the synchrony judgement task, measuring the ability to discriminate between synchronous and asynchronous stimuli, at theta (θ) in experiment 1, and at delta (δ), theta (θ) and alpha (α) frequencies in experiment 2.

(C) Accuracy (%) of selecting the correct movies that were presented with particular sounds, when the movies and sounds were flickering in synchrony (S) or out-of-synchrony (A; 90°, 180°, and 270° phase offsets combined), at theta (θ) in experiment 1, and at delta (δ), theta (θ) and alpha (α) frequencies in experiment 2.

(D) The difference in accuracy between the synchronous and asynchronous conditions at theta (θ) in experiment 1, and at delta (δ), theta (θ) and alpha (α) frequencies in experiment 2. Error bars represent 95% confidence intervals.

Figure 3. Results of the analysis of the phase differences between visual (red; MNI coordinates: 28, -96, -6) and auditory (blue; MNI coordinates: 64, -24, 6) sources at each phase offset. See also Figure S2.

(A) Maps of evoked power obtained from the visual-only and auditory-only unimodal conditions

370   projected onto visual and auditory cortices, showing visual (red) and auditory (blue) sources.

Evoked power was calculated as the difference between average evoked power between 0.75 and

2.75 s at each dipole in the auditory and visual unimodal conditions, contrasted with the average

evoked power in the same condition when trials were randomly assigned to each phase offset

condition (wherein the expected evoked power would be zero).

375   (B) Top: Amplitude normalised grand-averaged signals from the visual and auditory cortices

(lowpass filtered at 15 Hz) at 0° phase offset. The stimuli onset at time $= 0$. Bottom: Wrapped count

histogram of the instantaneous phase difference between the visual and auditory signals, for the

shaded time region (0.75-2.75 s). Inset: The direction and length of the mean resultant vector of the

phase differences.

380   (C-E), Same as (B), but for 90°, 180°, and 270° phase offsets, respectively.

385

390

STAR Methods

Experimental Model and Subject Details

400   Twenty-four healthy English-speaking young adults (mean age = 19.96 years, sd = 1.52; 15

female) completed the first experiment. One participant was left-handed. Most participants (20)

were granted experimental participation credit; the remaining participants received a small stipend

(£10.00) for their participation. The data from all participants were retained for the final analysis.

Ten participants were tested in the second experiment before an error in the program code was

405   discovered; the data from these participants were discarded, and 24 new healthy English-speaking

young adults (mean age = 22.92 years, sd = 4.26; 19 female) completed the second experiment. All

participants were right handed, and received a small stipend (£10.00) for their participation. The

data from all participants were retained for the final analysis.

Nine healthy English-speaking young adults (including the first author; mean age = 26.55

410   years, sd = 4.21; 3 female) completed the EEG manipulation check. One participant was left-handed.

The data from one additional participant was excluded from the final analysis due to technical

problems with the EEG recording.

In all experiments, informed consent was obtained for all participants.

415   *Method Details*

Apparatus

Tasks were programmed with MATLAB (R2013a; The Mathworks, Inc., Natick, MA, USA), using extensions from the Psychophysics Toolbox [32-34]. Visual stimuli were presented on a 21-inch CRT display with a screen refresh rate of 75 Hz, through an nVidia Quadro K600 graphics

420   card (875 MHz graphics clock, 1024 MB dedicated graphics memory; Nvidia, Santa Clara, CA, USA). Auditory stimuli were presented through Sennheiser HD 201 headphones (Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany) via a Sound Blaster Audigy 5/Rx audio card (Creative Technology Ltd., Singapore) in the two experiments, and through insert earphones (ER-3C; Etymotic Research, Elk Grove Village, IL) in the EEG manipulation check. The Psychophysics

425   Toolbox extension PsychPortAudio and ASIO4All (2.12; Steinberg Media Technologies GmbH) software were used to control the timing of the presentation of the auditory stimuli. A Pico ADC-212 oscilloscope (picotech.com) with a ThorLabs DET36A photodetector (thorlabs.de) and PicoScope for Windows (5.08.6) software were used to verify the phase offsets between the auditory and visual stimuli and the accuracy of movie and sound presentation timing. The

430   experiments were run from a solid-state hard drive on a Windows 7-based PC (3.40 GHz processor, 16 Gb RAM). A standard QWERTY computer keyboard to respond to the tasks.

Movies

The visual stimuli consisted of 96 (experiments 1 and 2) or 192 (EEG manipulation check)

435   movie clips of 3 second duration (76 frames total; framerate = 25). The movie clips were of emotionally neutral human activity (e.g., cars driving down a road, a ship being unloaded) or emotionally neutral activity in nature (e.g., a buffalo foraging for food, a turtle on a beach). For experiment 1 and the EEG manipulation check, all movies were luminance modulated with a 4 Hz sine wave, such that the movies visually flickered from no luminance to full luminance at 4 Hz. For

440   experiment 2, 32 of the 96 movies were luminance modulated with a 4 Hz sine wave, 32 different movies were luminance modulated with a 1.652 Hz sine wave, and the remaining 32 movies were luminance modulated with a 10.4721 Hz sine wave, such that movies visually flickered from no

luminance to full luminance at each frequency. In all experiments and conditions, all movies began

at 50% of full luminance, initially increasing in luminance. A sample movie is available in the

445   online content.

Sounds

The auditory stimuli consisted of 96 (192 in the EEG manipulation check) sound clips,

comprising 24 sounds from each of four instrument categories: acoustic guitar, electric guitar,

450   synthesiser, and an orchestra (eight instrument categories were used in the EEG manipulation

check). Sound clips were acquired from Apple Loops for Garage Band (6.0.5) and iLife Sound

Effects audio libraries freely available on most Apple computers, and from various unique

soundtracks, and were chosen such that amplitude modulation and rhythmic beats within the sounds

were minimal. Each sound clip was trimmed to a 3-second duration and pre-processed using

455   Audacity software (2.1.1; audacityteam.org). For each sound, a 100 ms linear amplitude fade-in and

fade-out was applied, and each channel was independently normalised to 0.0 dB maximum

amplitude. For experiment 1 and the EEG manipulation check, each of the sounds were amplitude

modulated with a 4 Hz sine wave (four times; once at each of 0°, 90°, 180°, and 270° phase offset

from the sine wave used to modulate the movie), such that the sounds 'flickered' from no volume to

460   full volume at 4 Hz. For experiment 2, each sound clip was amplitude modulated with a 4 Hz sine

wave, a 1.652 Hz sine wave, and a 10.4721 Hz sine wave, such that the sounds 'flickered' from no

volume to full volume at each of the three frequencies. Each sound clip was modulated with each of

the three sine waves four times: Once at each of 0°, 90°, 180°, and 270° phase offset from the sine

wave used to modulate the movie. A sample sound is available in the online content.

465

Procedure: Experiments 1 and 2

Participants sat in a well-lit testing room. Each experiment began with the completion of a

safety screening questionnaire and the provision of informed consent. Participants sat 60 cm from

the centre of the screen. A block of practice trials familiarised each participant with the experiment.

470 Following task instructions and completion of the practice trials, a curtain was drawn to separate the experimenter and the participant.

The associative episodic memory task consisted of six blocks, each of which consisted of an encoding phase, a distractor task, and an associative episodic memory test phase (Figure S1A). The order of block presentation was balanced so that each block of trials appeared in each serial position

475 an equal number of times across participants in each experiment. At the end of the memory blocks a perceptual synchrony task was carried out in order to test whether subjects were able to distinguish between the different phase offsets.

Encoding: In the encoding phase, participants were instructed to watch a series of sixteen movies, each of which were simultaneously presented with a sound clip (likened to a soundtrack for

480 the movie; Figure 1). During each trial, participants judged how well the sound suited the content of the movie, using the following example as a guide:

"Suppose the movie is of a gorilla sitting in some grass. The sun is up, and the gorilla is scratching itself and looking very content. Now suppose that the sound is some nice, soft, orchestral music… you can almost picture

485 David Attenborough narrating. In that case, you might say that the sound suits the content of the movie. On the other hand, if the sound was a distorted electric guitar, playing a wild solo, using a whammy bar and sounding crazy, then you might say that the sound does not suit the content of the movie very well."

490 These ratings were made using the number keys 1 (the sound does not suit the content of the movie at all) through 5 (the sound suits the content of the movie very well) on the keyboard. Four sounds from each of the four instrument categories were randomly paired with the 16 movies in each block of trials. Participants were told to remember the associations between the movies and the sounds, as there would be a memory test later. In experiment 1, four sounds (one modulated with a

495   wave at each phase offset with respect to the movie) from each of the four instrument categories were randomly paired with the 16 movies in each block of trials. Thus, each block of trials contained four trials at each phase offset between the movie and the sound. In experiment 2, two of the six blocks contained movies and sounds modulated at 4 Hz, two of the blocks contained movies and sounds modulated at 1.652 Hz, and two of the blocks contained movies and sounds modulated

500   at 10.472 Hz. Four sounds (two modulated with a wave with a 0° phase offset with respect to the movie, and two modulated with a wave with a non-zero phase offset with respect to the movie) from each of the four instrument categories were randomly paired with the 16 movies in each block of trials. Thus, each block of trials contained eight trials with a 0° phase offset between the movie and the sound, and eight trials with a non-zero phase offset between the movie and the sound. In all

505   experiments, the order of block presentation was balanced so that each block of trials appeared in each serial position an equal number of times.

Distractor: The distractor phase began after the presentation of the final movie and sound in each block. In the distractor phase, participants were instructed to count backwards by 3 from a random number presented on the screen (randomly drawn between 70 and 99 in experiment 1, and

510   randomly drawn between 180 and 199 in experiment 2, aloud, for 30 seconds.

Recall: The associative episodic memory test phase began after the distractor phase had been completed. In each trial, participants were presented with a screen that displayed still images from four of the movies that they had seen during the encoding phase of the same block. A sound clip that was presented with one of the four movies in the encoding phase was presented simultaneously.

515   On every trial, the four still images were from movies that were presented with a sound from the same instrument category. Participants selected the movie that was playing when they heard that sound during the encoding phase (Figure S1B).

Synchrony perception task: After completing all cycles of the associative episodic memory task, participants completed a synchrony judgement task; the instructions for this task were not

520   given until the participants had completed all cycles of the associative episodic memory task.

Participants were presented with a subset of 24 (six at each phase offset) of the 96 movie-sound pairings that they had seen throughout the blocks of the experiment, and were instructed to make a two-alternative forced-choice judgement as to whether the flicker of the luminance of the movie and the flicker of the volume of the sound were in-phase or out-of-phase. In experiment 1, the first 18 participants completed the task as described above; the final six participants also completed four additional synchrony judgement blocks, with the movies and sounds modulated at 1.652 Hz, 2.4721 Hz, 6.4721 Hz, or 10.4721 Hz in each block. These blocks served as pilot data for the second experiment. In experiment 2, participants completed three synchrony judgement blocks (one block at each of the three modulation frequencies), each containing 24 (six at each phase offset) of the 96 movie-sound pairings that they had seen throughout the blocks of the experiment.

Experiment 1 lasted approximately 40 minutes for the first 18 participants, and approximately 55 minutes for the final six participants. Experiment 2 lasted approximately 60 minutes, and the EEG manipulation check lasted approximately 120 minutes.

Procedure: EEG Manipulation Check

Participants sat in a well-lit testing room, 60 cm from the centre of the screen, and were prepared for EEG data collection. A block of practice trials familiarised each participant with the experiment. Following task instructions and completion of the practice trials, the experimenter monitored the participant via a webcam from an adjoining EEG recording control room.

The testing session was divided into 14 testing blocks. For the first twelve blocks, participants were instructed to watch a series of sixteen movies, each of which were simultaneously presented with a sound clip (likened to a soundtrack for the movie; Figure 1). During each trial, participants judged how well the sound suited the content of the movie, as in the other experiments. Four sounds from four of the eight instrument categories were randomly paired with the 16 movies in each block of trials. Each block of trials contained eight trials with a 0° phase offset, and eight trials with a 90°, 180°, or 270° phase offset between the movie and the sound, such that 96 of the 192 total trials had

a 0° phase offset between the movie and the sound, and the 90°, 180°, and 270° phase offset conditions consisted of 32 trials each.

The final two blocks consisted of a series of sound clips only (block 13) or movies only (block 550 14); each block contained 50 trials. In these blocks, participants were instructed to rate how pleasant the sound or the movie was, using the number keys 1 (the sound (movie) was very unpleasant) through 5 (the sound (movie) was very pleasant) on the keyboard.

EEG Methods

555 Continuous EEG data was recorded using a 128 channel BioSemi ActiveTwo system (BioSemi, Amsterdam, The Netherlands). Electrode positions were the 128 standard equidistant BioSemi sites; EOG generated from eye movements and blinks were recorded from three additional electrodes placed approximately 1 cm to the left of the left eye, 1 cm to the right of the right eye, and 1 cm below the left eye. Data were digitised using the BioSemi ActiView software, with a 560 sampling rate of 2048 Hz. Off-line analysis were performed with SPM12 (Wellcome Trust Centre for Neuroimaging) and FieldTrip [35] (version 2016031538).

EEG data were re-referenced to the average reference, and highpass filtered with a cutoff of 1 Hz (butterworth filter, zero phase, order 5), prior to resampling at 512 Hz. The data were then lowpass filtered with a cutoff of 20 Hz (butterworth filter, zero phase, order 5), and epoched (1000 565 ms before to 4000 ms after stimulus onset) and baseline corrected. Trials and channels with obvious artefacts (other than ocular artefacts that occurred between 3000 and 4000ms after stimulus onset) were removed by visual inspection, and by using Fieldtrip's visual artefact rejection procedure. Independent components analysis (ICA) was then applied to the data, and components related to ocular artefacts were removed. Any remaining artefacts were removed during a final visual 570 inspection of the data. On average, 22.72% of trails were removed from the data, and 5.4% of channels were removed. Sensor data were then interpolated (via triangulation of nearest neighbours)

and re-referenced to the average reference, if any channels had been removed during the artefact rejection procedure.

EEG electrode positions were measured on the first participant's head using a Fastrak

575 electromagnetic digitiser (Polhemus Inc., Colchester, VT, USA), and warped to match the orientation of the standard 10-5 electrode template provided in FieldTrip. Using the geometric locations of the electrodes and the template FieldTrip volume conduction and source models, sources were reconstructed over time and space using a linear constrained minimum variance (LCMV) beamformer. Given that activity in the auditory cortices have highly correlated time

580 courses with binaural stimulation, which poses a problem for accurate source localisation using a LCMV beamformer, we followed an established method that allows detection of correlated sources by calculating the surface Laplacian of the data and the leadfields to improve the spatial resolution of the data in the unimodal auditory and multimodal conditions [36].

In the unimodal conditions, time-frequency analysis was applied to each source ERP using

585 Morlet wavelet (width = 7) at our frequency of interest (4 Hz). Evoked power at each dipole was determined by averaging the evoked power between 0.75 and 2.75 seconds post stimulus onset. The evoked power at each dipole was contrasted with evoked power at each dipole in a randomised condition. For the randomised conditions, individual trials for each participant were randomly assigned a 0, 90, 180, or 120 phase offset by shifting the signal forward in time by 0, 32, 64, or 96

590 samples (0, 62.5, 125, or 187.5 ms). The expected value of the evoked power at relevant auditory and visual sources would thus be zero in the randomised conditions. For each participant, the evoked power difference at each source in each of the unimodal conditions was interpolated to the MNI 305 MRI template. The grand average of these power differences was interpolated onto the MNI 305 MRI template and displayed using MNI-space templates included with CARET (version

595 5.65; Washington University School of Medicine) for visualisation of sources of power greater than 90% of the maximum (Figure 3 and Figure S2). Coordinates for auditory and visual sources were

determined by finding the maximum power difference in regions corresponding to auditory and visual sensory areas.

In the multimodal conditions, we obtained common source filters separately. Source grand average EEG signals were calculated at the determined coordinates for the auditory and visual sources, and interpolated to the template structural MRI. Individual trial data were lowpass filtered with a cutoff of 15 Hz and grand averaged in each multimodal condition (0°, 90°, 180°, and 270°) at each determined source. The Hilbert transformation was applied to the grand averaged signals obtained at each determined source for each phase offset condition. The instantaneous phases were derived from the Hilbert transformed data, and unwrapped. The unwrapped instantaneous phase differences (modulus 2π) between the auditory source and the visual source was calculated between 0.75 sec and 2.75 sec in order to avoid effects of stimulus onset and offset, for each phase offset condition.

Quantification and Statistical Analysis

Effects of phase offset on memory performance were obtained by repeated measures ANOVA and/or pairwise t-tests (see main Results). In all cases except GLMMs, statistical significant was defined a $p < 0.05$. For GLMMs, a likelihood ratio of 5 or greater was defined as acceptably strong evidence in favour of a model.

Additional Analysis: Experiment 1

In addition to the analysis of variance on the proportion of correct responses in the different phase offset conditions, comparisons of generalised linear mixed models (GLMMs [37]) provided a metric of the strength of the evidence for the effects. Importantly, GLMMs allow specification of the distribution family of the effect of interest, thus allowing responses to be properly treated as binomially distributed data (i.e., using a logit regression). GLMMs compare models with a likelihood ratio, and here are corrected using Akaike's information criterion [38] to adjust for model

complexity. The corrected likelihood ratios are presented as 'bits (binary digits) of evidence' (i.e., on a log-base-2 scale). Royall [39] suggests that values greater than 5 represent strong evidence in favour of one model over the other.

The corrected log-likelihood ratio of 9.223 suggests strong evidence in favour of the 'unrestricted' model (that correct responses were a function of phase offset), compared to the 'restricted' model (that correct responses were not a function of phase offset). Z- and P- values for the logit regression coefficients support the pairwise T-tests presented in the main text: the 90°, 180°, and 270° phase offset conditions were each significantly different from the 0° phase offset condition: $Z = -2.238$, $P = 0.025$; $Z = -2.650$, $P < 0.01$; and $Z = -3.180$, $P < 0.01$, respectively.

Analyses of discriminability (d-prime) and response bias (beta) on responses in the synchrony judgement task revealed no significant effect of discriminability (H_0: mean $= 0$: $T(1,23) = 1.224$, $P = 0.233$), or response bias (H_0: mean $= 1$: $T(1,23) = -0.174$, $P = 0.864$). One participant had a 100% hit rate, so for the purposes of the d-prime and beta calculations, one hit was substituted for a miss for this participant. Thus, perceptual judgement of synchrony was no better for synchronous or asynchronous conditions, suggesting that perceptual judgements of synchrony were not responsible for performance on the episodic memory task.

In order to rule out the potentially confounding effect of audio-visual synchrony on the rating judgements of how well the sound suited the content of the movie, an additional repeated measures ANOVA was performed, and revealed no effect of phase offset on ratings of how well the sounds suited the movies: $F(3,69) = 1.970$, $P = 0.125$, suggesting that it was not the case that participants rated the sound as suiting the movie differently in the different phase offset conditions; thus the suitability ratings were not influenced by phase offset, and do not explain performance on the episodic memory task.

As a final control analysis to rule out any effect of the rating judgements of how well the sound suited the content of the movie on memory performance, we equalised the number of trials in each of the four phase offsets at each value on the rating scale (removing 256 trials, leaving 2048 trials

for the present analysis). A repeated measures ANOVA revealed the effect of phase offset on

650 memory performance: $F(3,69) = 5.155$, $P < 0.01$, again confirming our earlier result.

Additional Analysis: Experiment 2

GLMMs support the results of the ANOVAs. The corrected log-likelihood ratio of 5.62

suggests strong evidence in favour of the model including a term capturing synchrony (i.e., that

655 correct responses were a function of phase offset) in the theta condition, but not in the delta (-2.88)

or alpha (-2.85) conditions.

Analyses of discriminability (d-prime) and response bias (beta) in the synchrony judgement

task revealed no significant effect of discriminability in the theta condition (H_0: mean $= 0$: $T(1,23) =$

1.269, $P = 0.217$), or response bias (H_0: mean $= 1$: $T(1,23) = 0.143$, $P = 0.888$), or the alpha

660 condition (H_0: mean $= 0$: $T(1,23) = 0.002$, $P = 0.999$), or response bias (H_0: mean $= 1$: $T(1,23) = -$

0.3062, $P = 0.762$). In the delta condition, there was a significant effect of discriminability (H_0:

mean $= 0$: $T(1,23) = 3.059$, $P = 0.006$), and response bias (H_0: mean $= 1$: $T(1,23) = -4.642$, $P <$

0.001). Synchrony could be more easily detected in the delta condition, importantly this did not lead

to improved performance on the episodic memory task for synchronous stimuli. In the alpha and

665 theta conditions, the perceptual judgement of synchrony was no better for synchronous or

asynchronous conditions. Overall, there is no suggestion that perceptual judgements of synchrony

were responsible for performance on the episodic memory task in these conditions.

As additional control analyses, participants were split into two groups based on their

performance (better or worse than the median performance) on the synchrony judgement task in

670 each frequency condition. A 2 (group) by 2 (synchrony) repeated measured ANOVA for the theta

condition revealed no effect of group: $F(1,22) = 0.226$, $P = 0.639$, and no interaction between group

and synchrony: $F(1,22) = 0.003$, $P = 0.960$ on memory performance. The effect of synchrony

remained significant: $F(3,22) = 10.663$, $P = 0.004$. The 2 by 2 repeated measures ANOVA for the

delta condition revealed no effect of group: $F(1,22) = 0.186$, $P = 0.671$, and no interaction between

675 group and synchrony: $F(1,22) = 0.217, P = 0.646$ on memory performance. The lack of the effect of synchrony remained unchanged: $F(1,22) = 0.027, P = 0.871$. The ANOVA for the alpha condition revealed no effect of group: $F(1,22) = 2.406, P = 0.135$, and no interaction between group and synchrony: $F(1,22) = 0.220, P = 0.644$ on memory performance. The lack of the effect of synchrony remained unchanged: $F(1,22) = 0.005, P = 0.946$. All ANOVAs suggest that perceptual judgements

680 of synchrony were not responsible for performance on the episodic memory task.

In order to rule out the potentially confounding effect of audio-visual synchrony on the rating judgements of how well the sound suited the content of the movie, an additional 3 (frequency) by 2 (synchrony) repeated measures ANOVA was performed, revealing a significant interaction between frequency and synchrony: $F(2,46) = 15.396, P < 0.001$. At the alpha frequency the sound not was

685 judged as suiting the content of the movie differently in the synchronous versus the asynchronous condition: $T(1,23) = 0.418, P = 0.680$. In the delta frequency condition, the sound was judged as suiting the content of the movie better in the synchronous compared to the asynchronous condition: $T(1,23) = 3.075, P = 0.005$. However, unexpectedly, in the theta condition the sound was judged as suiting the content of the movie better in the asynchronous compared to the synchronous condition:

690 $T(1,23) = -4.668, P < 0.001$. Thus, while in the second experiment the suitability ratings were influenced by synchrony in the delta and theta conditions, the direction of the effect cannot explain differential performance on the episodic memory task in the theta condition, and the lack of differential performance in the delta condition.

As a final control analysis to rule out any effect of the rating judgements of how well the sound

695 suited the content of the movie on memory performance, we equalised the number of trials in each of the two synchrony conditions at each of the three frequency conditions, at each value on the rating scale (removing 208 trials, leaving 2096 trials for the present analysis). T-tests confirmed the effect of synchrony at the theta frequency, even when trial numbers were equalised across the five ratings: $T(1,23) = 2.484, P = 0.021$. The lack of effect at the alpha and delta frequencies remained

700 unchanged: $T(1,23) = 0.390, P = 0.700$, and $T(1,23) = -0.162, P = 0.873$, respectively. Furthermore,

the effect at theta remained significantly different from the combined effects of delta and alpha: $T(1,23) = 1.823$, $P = 0.040$, again confirming our earlier result.

Control Experiment

705   Improved binding at the perceptual level for synchronously flickering stimuli, relative to asynchronously flickering stimuli, could explain the pattern of results observed in our experiments, if improved binding at a perceptual level results in improved associative memory formation. We performed a control experiment wherein participants were presented with the same stimuli used in experiment 1 (sounds and videos modulated with a 4 Hz sine wave), or unmodulated pairs of

710   sounds and videos, in separate blocks. Two blocks of 16 trials contained sounds and videos modulated as in experiment 1, but restricted to 0-degrees or 180-degrees phase offset (8 trials at each offset in each block). Two blocks of 16 trials contained sounds and videos that were presented at 100% amplitude or luminance, respectively. In order to control for the amount of information presented between the conditions, the unmodulated sounds and videos were trimmed to half of their

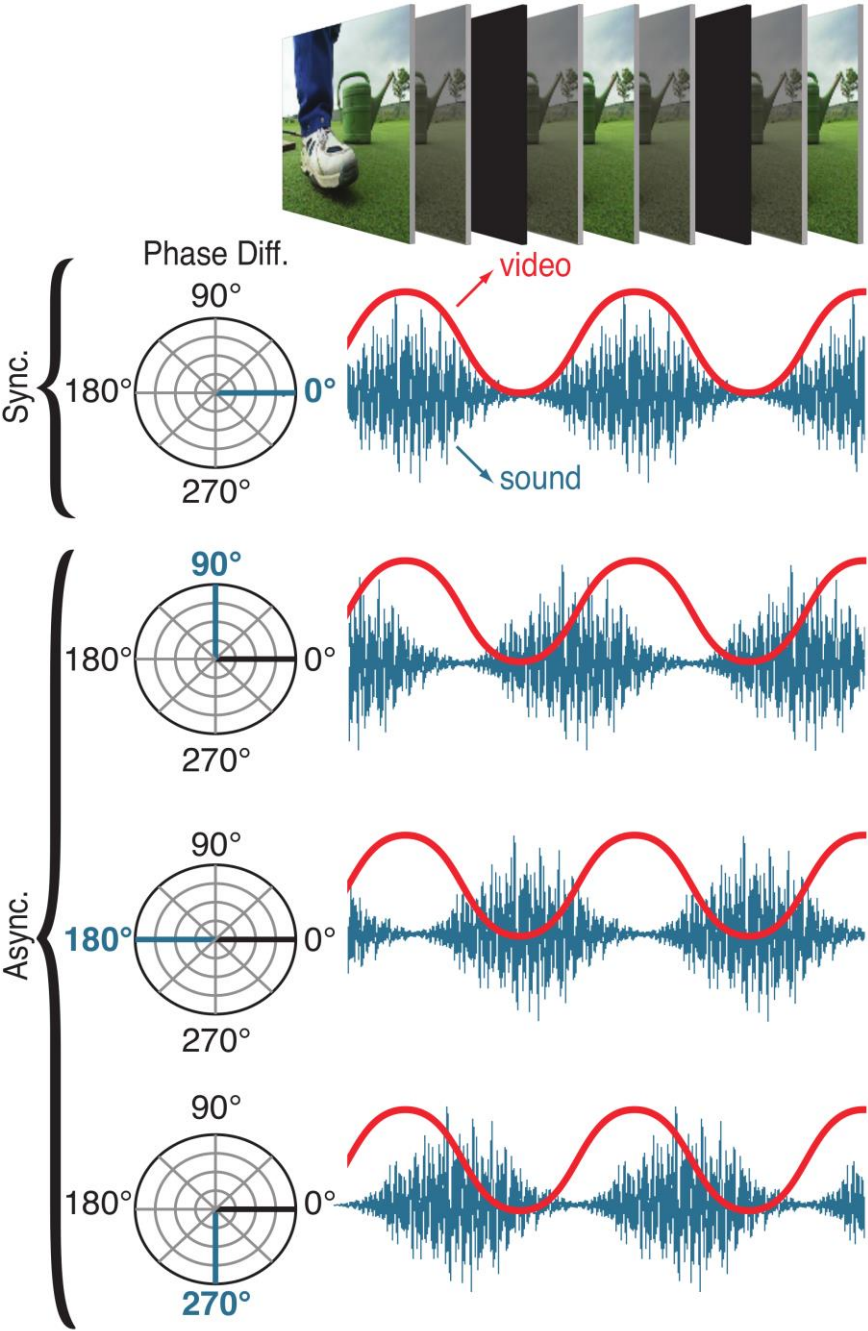715   length (i.e., 1.5 second unmodulated sounds and videos).

Twenty-four healthy English-speaking young adults (mean age = 20.1 years, sd = 1.21; 19 female) completed the control experiment. All participants were right-handed. Participants were granted experimental participation credit or a small stipend (£10.00) for their participation. The data from all participants were retained for the analysis.
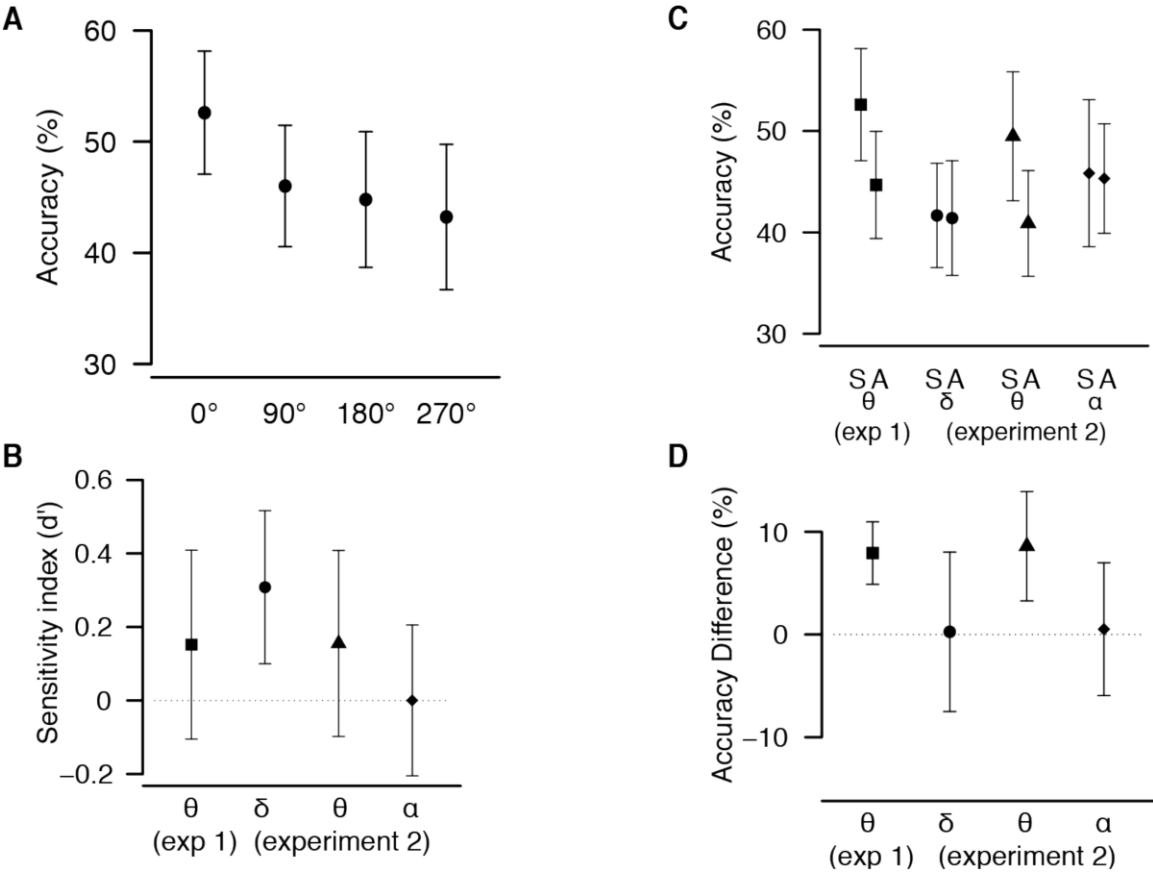
720   Memory performance and differences between the 0-degree phase offset, 180-degree phase offset, and the unmodulated stimuli are shown in Figure S3. Analysis of variance (ANOVA) reveals a main effect of condition: $F(2,46) = 9.48$, $P < 0.01$. Memory performance was best in the 0-degree phase offset condition (Figure S3). A pairwise t-test reveals a significant difference between the 0-degree and 180-degree phase offset conditions: $T(1,23) = 4.01$, $P < 0.01$. A pairwise t-test between

725   0-degree and unmodulated stimuli also reveals a significant difference: $T(1,23) = 3.24$, $P < 0.01$.
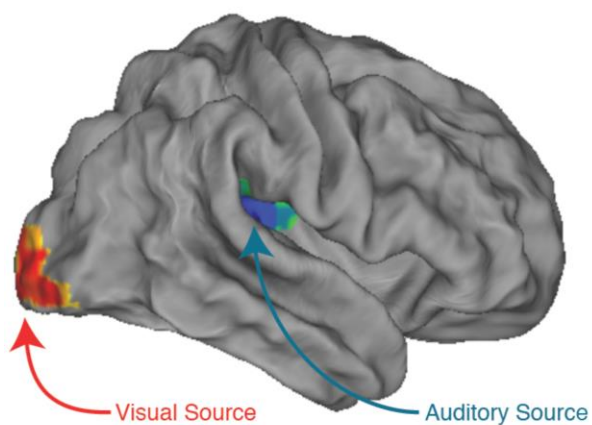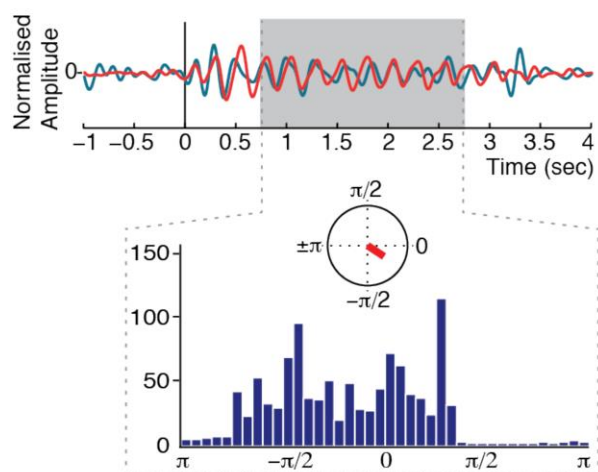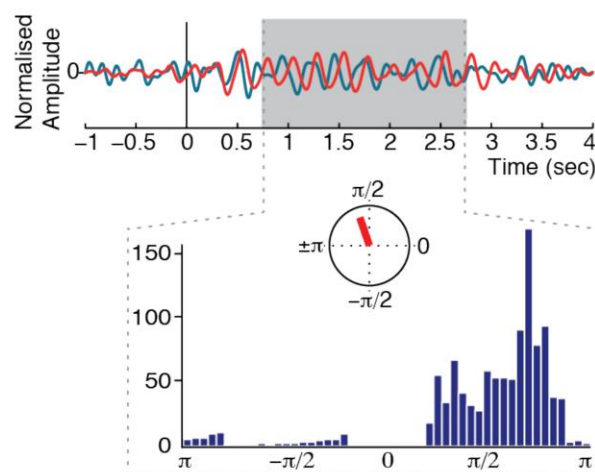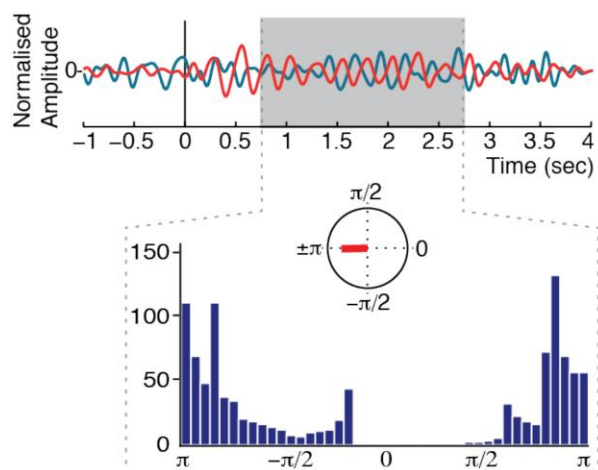
However, memory for the unmodulated stimuli was not different from the stimuli presented at 180-degrees phase offset: $T(1,23) = -0.92, P = 0.37$.

Synchronous stimulus presentation at theta improved associative memory, relative to both asynchronous presentation and non-flickering (unmodulated) stimuli. This result is surprising, given that the flickering stimuli were distracting, whereas the unmodulated stimuli were presented in a state ideal for perception. Thus, our results suggest that the theta-induced memory effect is indeed a memory effect, and not an effect resulting from binding at the pereptual level, before associative memory processes take place.

Figure 1

760   **Figure 2**

Figure 3