Current Biology

Report

Altering stimulus timing via fast rhythmic sensory stimulation induces STDP-like recall performance in human episodic memory

Highlights

- Multimodal input timing can be manipulated by 37.5 Hz rhythmic sensory stimulation
- Episodic recall is enhanced when a corresponding sensory modality leads by 6.67 ms
- When the corresponding sensory modality is delayed by 6.67 ms, recall is impaired
- This memory effect is consistent with the simulation of an STDP computational model

Authors

Danying Wang, Kimron L. Shapiro, Simon Hanslmayr

Correspondence

danying.wang@glasgow.ac.uk (D.W.),
simon.hanslmayr@glasgow.ac.uk (S.H.)

In brief

Wang et al. use rhythmic sensory stimulation at 37.5 Hz to control visual and auditory input timing on the order of ms. Episodic cued recall is enhanced if the corresponding modality leads by 6.67 ms. Recall is impaired when the modality leads by 20 ms. This memory effect is consistent with the results simulated by a computational model of STDP.

Wang et al., 2023, Current Biology *33*, 3279–3288
August 7, 2023 © 2023 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2023.06.062

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

Altering stimulus timing via fast rhythmic sensory stimulation induces STDP-like recall performance in human episodic memory

Danying Wang,[1,2,*] Kimron L. Shapiro,[2] and Simon Hanslmayr[1,2,3,4,*]
[1]School for Psychology and Neuroscience and Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow G12 8QB, UK
[2]School of Psychology and Centre for Human Brain Health, University of Birmingham, Birmingham B15 2TT, UK
[3]Twitter: @SimonHanslmayr
[4]Lead contact
*Correspondence: danying.wang@glasgow.ac.uk (D.W.), simon.hanslmayr@glasgow.ac.uk (S.H.)
https://doi.org/10.1016/j.cub.2023.06.062

SUMMARY

Episodic memory provides humans with the ability to mentally travel back to the past,[1] where experiences typically involve associations between multimodal information. Forming a memory of the association is thought to be dependent on modification of synaptic connectivity.[2,3] Animal studies suggest that the strength of synaptic modification depends on spike timing between pre- and post-synaptic neurons on the order of tens of milliseconds, which is termed "spike-timing-dependent plasticity" (STDP).[4] Evidence found in human *in vitro* studies suggests different temporal scales in long-term potentiation (LTP) and depression (LTD), compared with the critical time window of STDP in animals.[5,6] In the healthy human brain, STDP-like effects have been shown in the motor cortex, visual perception, and face identity recognition.[7–13] However, evidence in human episodic memory is lacking. We investigated this using rhythmic sensory stimulation to drive visual and auditory cortices at 37.5 Hz with four phase offsets. Visual relative to auditory cued recall accuracy was significantly enhanced in the 90° condition when the visual stimulus led at the shortest delay (6.67 ms). This pattern was reversed in the 270° condition when the auditory stimulus led at the shortest delay. Within cue modality, recall was enhanced when a stimulus of the corresponding modality led the shortest delay (6.67 ms) compared with the longest delay (20 ms). Our findings provide evidence for STDP in human episodic memory, which builds an important bridge from *in vitro* studies in animals to human memory behavior.

RESULTS

Donald Hebb proposed synaptic plasticity as the neuronal basis for learning and memory in his seminal book,[14] writing, "Neurons that fire together wire together." In line with this postulate, spike-timing-dependent plasticity (STDP) that requires two neurons firing within a time window on the order of tens of milliseconds to induce a synaptic change has been found[4,15] (Figure 1A). A recent study using a rhythmic sensory entrainment (RSE) approach[16] precisely controlled the input phase offsets in theta (4 Hz) to demonstrate a role for theta-phase-mediated plasticity in human episodic memory.[17] However, given that the STDP occurs on a much faster timescale, it is unknown if human episodic memory is influenced by the inputs' relative timing in the order of tens of milliseconds. In this study, we examined this using RSE at 37.5 Hz, which allowed altering the relative timing between visual and auditory stimuli at fine temporal resolution with four phase offset conditions: 0°, 90°, 180°, and 270° (Figure 1B). Accordingly, the visual inputs precede the auditory inputs by 0, 6.67, 13.33, or 20 ms, respectively. The corollary auditory inputs precede the visual inputs by 0, 20, 13.33, or 6.67 ms, respectively. Participants were asked to memorize the pairs of video and sound clips. During recall, we cued participants' memory with each stimulus modality in a between design. Group 1 participants were cued with a video and asked to recall the paired sound. Group 2 was cued with a sound and asked to recall the paired video (Figure 1C). Our main prediction was that recall accuracy decreases or increases as a function of (1) the phase offset between auditory and visual stimuli and (2) the modality of the memory cue.

Simulating the results with an STDP computational model

To formalize the predictions for our results, we simulated the paradigm with a computational model that implements the STDP learning rule. Two groups of neurons that are simulated by an integrate-and-fire equation (cf. Parish et al.[18] and Wang et al.[19]) received two stimuli. The stimuli were modulated at 37.5 Hz with four phase offsets: 0°, 90°, 180°, and 270°. Synaptic weights from one group of neurons (e.g., visual) to the other group (auditory) were rewarded if visual neurons fired before auditory neurons and punished if auditory neurons fired first. The weight changes decayed exponentially over time.[20]

To evaluate model recall performance, weights from the visual group to the auditory group and from the auditory group to the visual group were averaged across 192 simulations for each

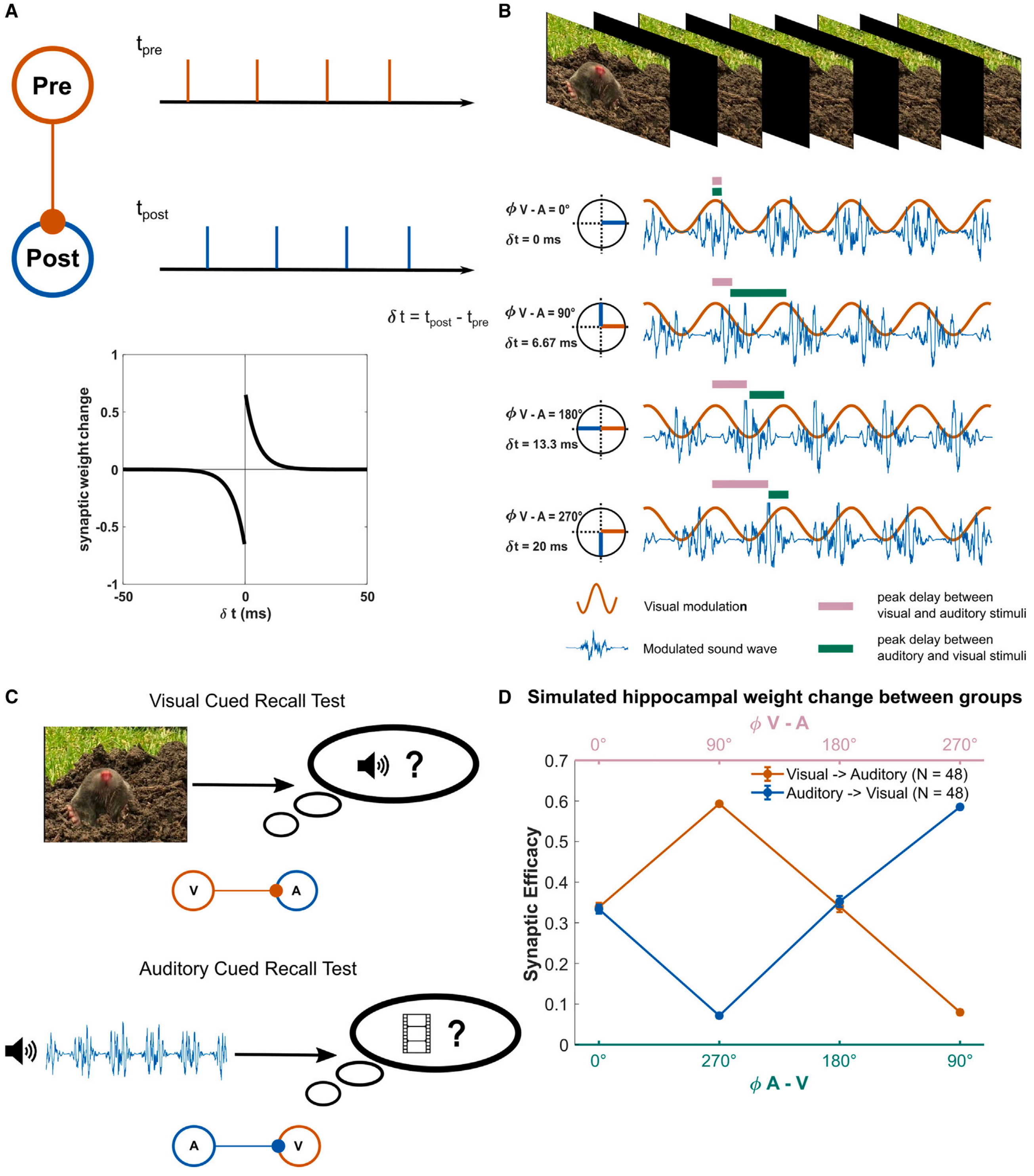


Figure 1. Experiment design

(A) Schematic of the STDP framework. Synaptic modification depends on spike timing between a pre-synaptic neuron and a post-synaptic neuron.

(B) The encoding phase involved viewing a 3 s video and listening to a 3 s sound clip. The luminance and amplitude of videos and sounds were modulated at 37.5 Hz with four phase offsets, 0°, 90°, 180°, and 270°. The pink bars represent peak delays between a video and a sound. The teal bars represent peak delays between a sound and a video.

(legend continued on next page)

phase offset condition. The model reveals that the weights from the visual group to the auditory group were higher compared with the weights from the auditory group to the visual group when the visual stimulus led the auditory stimulus by 90° (6.67 ms). This pattern was reversed when the visual stimulus led by 270° (20 ms), which means that the auditory stimulus led the visual stimulus by 90° (Figure 1D). The weights did not differ between 0° and 180° conditions. This is because half the pre-synaptic spikes precede the post-synaptic spikes, whereas half the pre-synaptic spikes follow the post-synaptic spikes, thus causing reward and punishment on weights to cancel each other. Therefore, we predicted the strongest memory differences to occur between the 90° and 270° conditions, depending on the modality of the cue and the target.

Specifically, visually cued recall accuracy should be higher than auditorily cued recall accuracy when the visual stimulus led the auditory stimulus by 90°. This pattern should be reversed in the condition where the visual stimulus led by 270° (i.e., when the auditory stimulus led the visual stimulus by 90°). Moreover, visually cued recall accuracy should be better when the visual stimulus led the auditory stimulus by 90° compared with 270°, whereas the pattern should be reversed for auditorily cued recall accuracy. Given the directionality of synaptic modification caused by the temporal order in STDP,[21] the statistical comparisons between modalities and phase offset conditions used one-tailed t tests if not specified.

Recall accuracy as a function of the actual phase difference between entrained visual and auditory activity

EEG was recorded for 24 participants in each group, which allowed confirmation of corresponding sensory modulation at 37.5 Hz at the specified phase offsets (Figure S1). Importantly, the difference between visual and auditory transduction delays causes the auditory domain to reach the cortex approximately 40 ms before the visual domain.[22–24] Therefore, we added a 40 ms delay before the auditory stimulus onset to approximate simultaneous processing in visual and auditory regions.[17,19] However, because of the 37.5 Hz modulation frequency, just a few milliseconds difference in the transduction delay would be detrimental for phase modulation (e.g., 5 ms corresponds to 67.6° at 37.5 Hz).

To compensate for this problem, we computed the actual phase differences between visual and auditory regions in each experimental condition to label each condition. Source activity was reconstructed from each region of interest (ROI) (Figure 2A). The event-related potential (ERP) was computed for each condition and each ROI. Figure 2B reveals that the actual instantaneous phase differences between visual and auditory grand average ERPs (N = 48) were 180° off from the expected phase offset conditions. The mean direction was statistically confirmed to be 0°, 90°, 180°, and 270° in the experimental conditions 180°, 270°, 0°, and 90°, respectively (V test; all p values = 0).

We relabeled our experimental conditions based on the actual phase offsets before investigating recall accuracy. Recall accuracy in each condition for each participant was normalized by their mean performance (Figure 3A). An ANOVA with the repeated-measures factor "phase offset condition" (90° versus 270°) and a between-subject factor "cue condition" (visual versus auditory) did not reveal an interaction ($F(1, 90) = 2.323$, $p = 0.131$). Consistent with our STDP model, visually cued recall accuracy in the actual 90° condition was higher than the auditorily cued recall accuracy. An independent-samples t test confirmed this difference to be statistically significant ($t(90) = 2.330$, $p = 0.011$, Cohen's $d = 0.486$). However, the auditorily cued recall accuracy did not differ from the visually cued recall accuracy in the 270° condition ($t(90) = -0.103$, $p = 0.459$).

To investigate if the results were caused by the individual variability in entrainment strength, we computed the inter-trial phase coherence (ITPC) at the stimulation frequency 37.5 Hz (Figures 3B and S1). An ANOVA with the repeated-measures factor phase offset condition (90° versus 270°) and factor "brain region" (visual versus auditory) indicated a significant interaction on the mean ITPC ($F(1,46) = 19.726$, $p < 0.001$, $\eta^2p = 0.3$). In the visual cortex, ITPC in the 90° condition in which the visual stimulus led was significantly stronger than in the 270° condition ($t(47) = 3.755$, $p < 0.001$, Cohen's $d = 0.542$), whereas in the auditory cortex, ITPC in the 90° condition was significantly weaker than in the 270° condition ($t(47) = -2.733$, $p = 0.004$, Cohen's $d = -0.395$; Figure 3B). Moreover, the advantage of the ITPC in the realigned 90° condition in the corresponding sensory region was positively correlated with the recall advantage in the 90° condition in the corresponding cue group ($r = 0.308$, $p = 0.033$, two-tailed; Figure 3C). The individual variability in sensory entrainment was linked with the recall accuracy advantage of the shortest delay led by the corresponding sensory modality.

Recall accuracy as a function of single-trial auditory and visual phase difference

Notably, Wang et al.[25] observed considerable trial-by-trial variation of phase differences arising between sensory cortices even though the phase difference of the sensory stimulation was constant. Therefore, we investigated if such a variance contributed to the discrepancy between the behavioral results and the model predictions. Instantaneous phase differences between band-pass filtered (35 and 40 Hz) visual and auditory activity were averaged between 0.5 and 2.5 s for each trial. Based on this value, single trials were sorted and divided into four equally sized bins (Figure S2). This procedure realigned each trial to the actual onset of sensory responses to the external stimuli, thus reducing any possible factors that cause a variation of phase differences. As a sanity check, we ensured that the resulting grand average ERP in each phase bin showed phase concentration at the intended directions, as confirmed by V test (N = 48; Figure 4A).

(C) Visually cued recall presented the video in the memory test phase. Participants were asked to select the correct sound. In the auditorily cued recall experiment, participants were cued with the sound and asked to recall the paired video.

(D) Hippocampal weight changes between two groups of neurons as a function of phase offset conditions, simulated by the STDP computational model. The inputs were modulated by a 37.5 Hz sine wave with the same phase offsets as in the experiments. The weights were averaged across 192 simulations for each condition. Error bars represent SE.

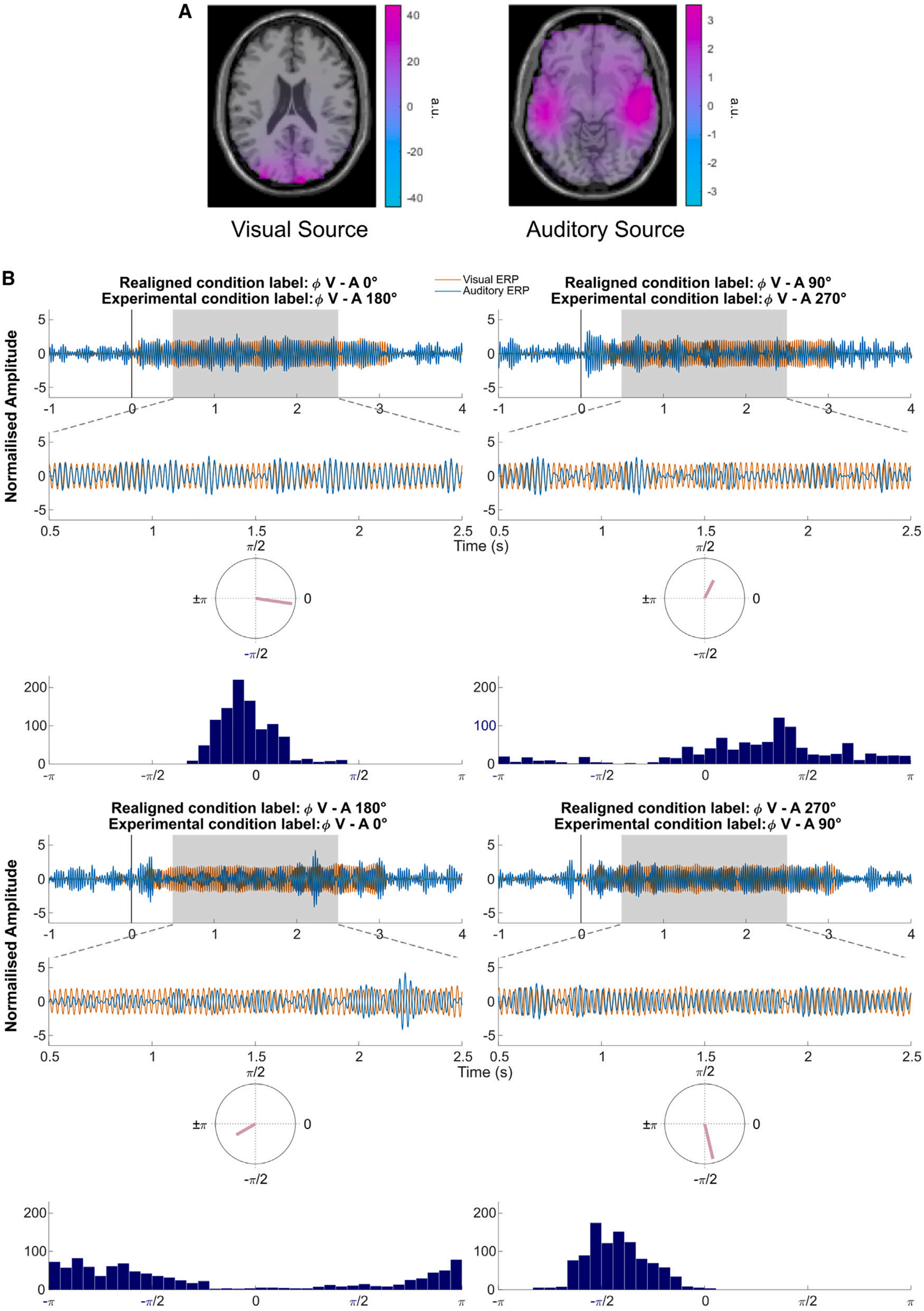


Figure 2. Phase differences between visual and auditory grand average ERP in each phase offset condition

(A) Source localization of visual and auditory sources in the unimodal conditions. Visual source, MNI coordinates of regions of interest (ROIs): 10, −99, and 20. Auditory sources, MNI coordinates of ROIs: right, 50, −19, and −10; left, −50, −31, and 0. Unimodal stimuli were modulated at 4 Hz. Evoked power was averaged

(legend continued on next page)

After single-trial sorting, the proportion of remembered trials in each phase bin was calculated (Figure 4B). Importantly, the resulting pattern resembles the STDP model-simulated data. The normalized recall score in phase bin 2 (i.e., 90°) was higher for the visually cued group compared with the auditorily cued group. Remarkably, this pattern was reversed in phase bin 4 (i.e., 270°) where the auditorily cued recall was better. An ANOVA with the repeated-measures factor “phase bin” (2 versus 4) and a between-subject factor cue condition (visual versus auditory) indicated a significant interaction ($F(1,44) = 7.055$, $p = 0.011$, $\eta^2 p = 0.138$). Independent-samples t tests confirmed that the normalized recall score in the visually cued group was significantly better than in the auditorily cued group in bin 2 ($t(44) = 2.384$, $p = 0.011$, Cohen’s $d = 0.703$), whereas this pattern was reversed in bin 4 ($t(44) = -1.971$, $p = 0.027$, Cohen’s $d = -0.581$). Furthermore, a paired-samples t test confirmed the hypothesis that visually cued recall was better when the visual stimulus led by the shortest delay, i.e., 90° or 6.67 ms (bin 2), relative to the longest delay, i.e., 270° or 20 ms (bin 4) ($t(22) = 1.931$, $p = 0.033$, Cohen’s $d = 0.403$). This pattern was reversed in the auditory cue condition. That is, recall accuracy in bin 4 was higher compared with bin 2 ($t(22) = -1.842$, $p = 0.039$, Cohen’s $d = -0.384$). Together, these results suggest that trial-by-trial phase differences between rhythmically stimulated visual and auditory sources generate memories that are consistent with the STDP model-simulated results.

Link between the simulated data and the empirical data in the hippocampus

The computational model implicated that STDP learning happens in the hippocampus. To investigate if the hippocampus is involved in the empirical data, ITPC at 37.5 Hz was computed at the whole-brain level for the encoding trials that were subsequently remembered and forgotten. A cluster-based permutation test restricted to the hippocampus confirmed a significant difference in the ITPC between remembered trials and forgotten trials localized in the hippocampus ($p_{corr} < 0.05$; Figure 4C). No significance was shown when the same analysis was applied to the prefrontal cortex (PFC). To link these results to the computational model, we split the simulated trials into high and low synaptic weight change trials, thus resembling later remembered and later forgotten trials, respectively. Consistent with the empirical results, the simulated data showed stronger ITPC for the high weights trials compared with low weights trials (Figures 4D and 4E). This consistency between the empirical data in the hippocampus and the simulated data suggests a hippocampal involvement in the STDP-like memory effect.

DISCUSSION

Forming lasting associations on a one-shot basis, which is the hallmark of episodic memory, is thought to depend on synaptic plasticity. STDP is one of the best documented mechanisms in animals and emphasizes that the temporal order and interval between two spikes determine the direction and strength of synaptic modification. Our results are the first to reveal that stimulus timing on the order of milliseconds influences human episodic memory, which is consistent with STDP. Previous studies show that *in vitro* human synapses follow the STDP rule, although the time window for inducing long-term potentiation (LTP) is wider than the classic rule observed in juvenile rats.[4–6,21,26] We simulated our experimental procedure with a model that implements the classic STDP rule.[18,19] The recall accuracy resembled the pattern of the model simulated synaptic weight changes, which suggests that the classic rule is sufficient to account for our results. Consistent with the human *in vivo* studies in perception,[9,11–13] our results demonstrate that near-synchronous cross-modality stimulus presentation enhances or impairs episodic memory association, depending on which modality is leading.

Using a related memory paradigm, previous studies[17,25] showed that synchronous presentation at theta frequency enhances episodic memory formation. However, the temporal resolution is considerably lower with 4 Hz modulated stimuli. We speculate that spike timing might be coordinated to interact with the theta-phase-dependent learning mechanism.[19,27,28] Indeed, memory performance was significantly worse for those conditions where the peak delays were outside of the STDP window. Given that memory was cued only unidirectionally, we could not experimentally disentangle the role of STDP, which, in turn, underscores the importance of this study.

Interestingly, we found a modality preference for the presentation order and interval at corresponding sensory cortices, as suggested by the ITPC at the stimulation frequency 37.5 Hz. The ITPC in the 90° condition was stronger than in the 270° in the corresponding sensory region if the stimulus of a modality led, which is consistent with previous findings on cortical STDP, suggesting a directionality-specific excitability in the corresponding sensory cortices.[29,30] Behaviorally, the advantage of the shortest delay led by the stimulus of a modality was linked with the corresponding cortical response of the modality, which might suggest why the large sample did not fully resemble the simulated STDP pattern. The individuals without the recall advantage in the 90° condition might have had weaker cortical responses indexed by the ITPC differences. This STDP-like cortical response can guide to realign the experimental conditions without computing the instantaneous phase differences between the sensory regions, thus unmasking the behavioral STDP-like effects. Similarly, when behavioral measurement is difficult, the STDP-like ITPC effects can be a proxy for the behavioral effects, which provides a complementary tool to indicate synaptic modification of human episodic memory.

between 3.5 and 4.5 Hz, and between 0.75 and 2.75 s at each virtual electrode in the unimodal visual and auditory conditions. The values (in arbitrary units) were normalized by averaged evoked power of pseudo baseline conditions, which the trials were selected randomly to shift by 0°, 90°, 180°, or 270° (STAR Methods).
(B) Phase differences between visual and auditory sources in each phase offset condition. Grand average ERP signals (N = 48) at the ROIs were band-pass filtered at 35–40 Hz. Amplitude was normalized. Instantaneous phase differences were averaged between 0.5 and 2.5 s (shaded time window) and plotted by the mean resultant vector on a unit circle. Histograms showed wrapped instantaneous phase differences between 0.5 and 2.5 s. See also Figures S1A and S1B for the time frequency representations (TFRs) of ITPC in each condition at each ROI.

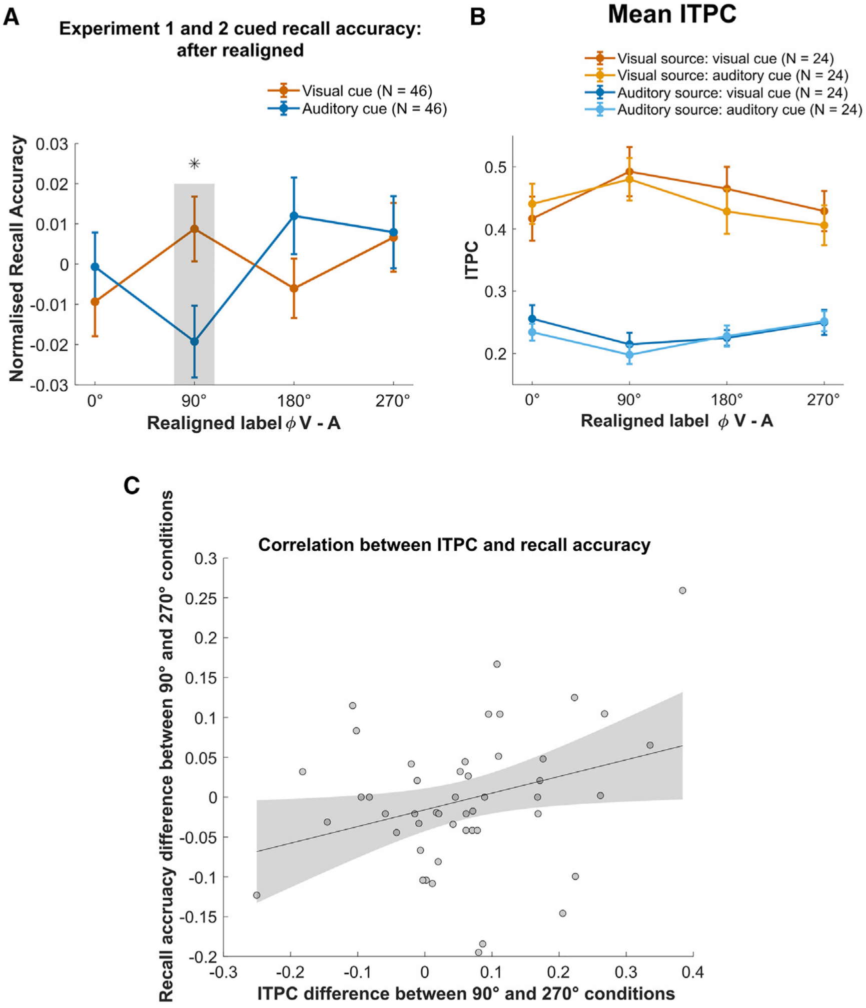


Figure 3. Behavioral results based on the actual phase difference between visual and auditory grand average ERP

(A) Recall accuracy (normalized by subtracting the mean across phase offset conditions) in each realigned phase offset condition based on the actual phase differences between visual and auditory grand average ERP. Error bars represent SE. *$p < 0.05$. N = 46 for each group. See also Figure S3A for the results of raw recall accuracy.

(B) Mean ITPC that was averaged between 0.5 and 2.5 s and between 37 and 38 Hz for each cue group (N = 24 for both groups) and each realigned phase offset condition at each source. Error bars represent SE. See also Figure S1C for the ITPC in each condition at each ROI as a function of frequency.

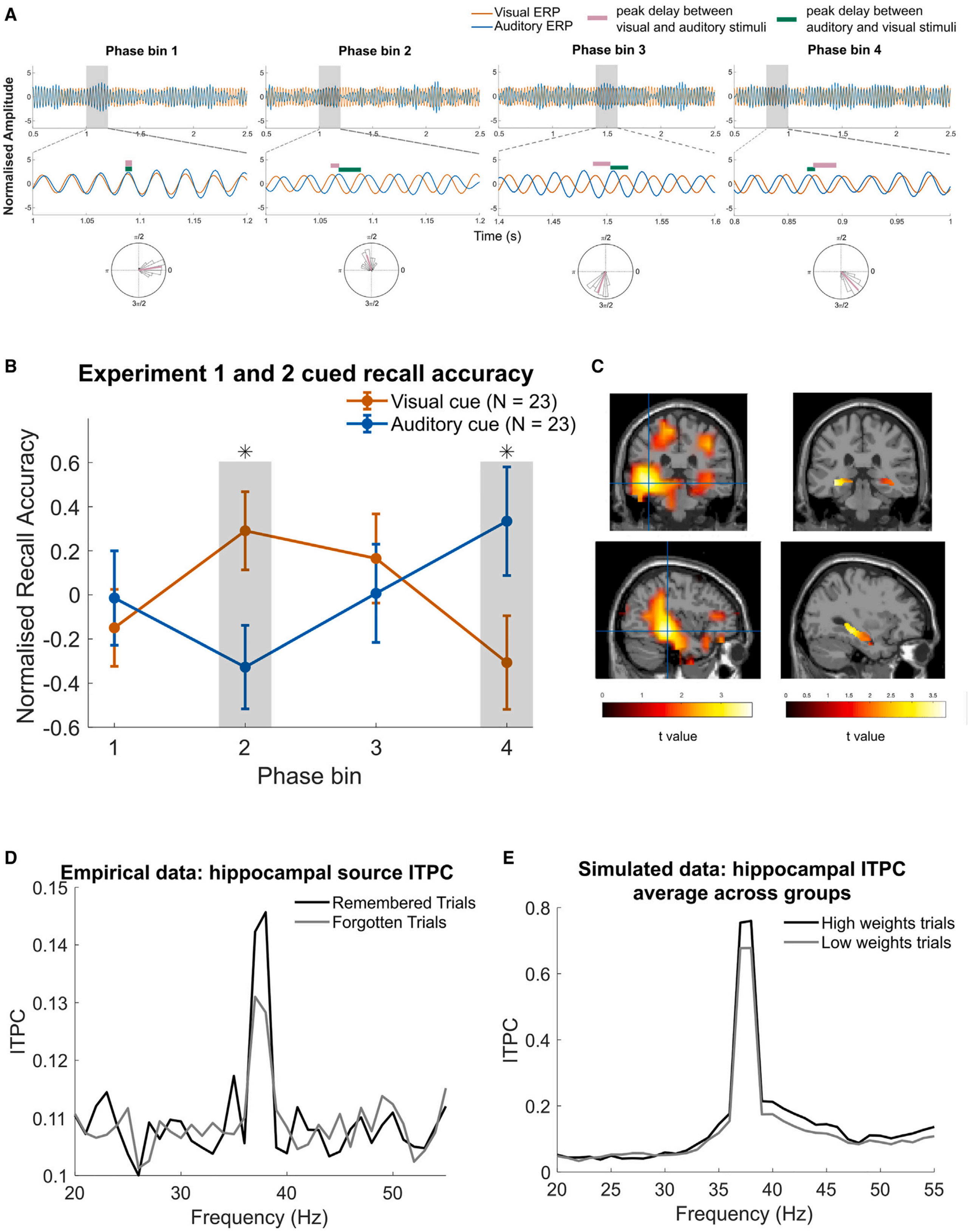
(C) Larger ITPC difference in the corresponding sensory regions was linked with higher recall accuracy difference in the corresponding cue group between realigned 90° and 270° conditions. Shading areas represent 95% confidence bound.

Neuronal co-firing in the medial temporal lobe at short delays predicted successful memory, whereas longer delays predicted memory failure, which is consistent with STDP.[31] Near-synchronous firing of hippocampal neurons leads to effective synaptic connectivity, thus supporting the formation of associations. Consistently, studies in rodents and humans have demonstrated that 40 Hz multisensory stimulation can reach the hippocampus, also improving the hippocampal function and, in turn, improving cognitive function[32–34] (but see Schneider et al.[35] and Soula et al.[36]). Therefore, the hippocampus may be responsible for the STDP-like memory effect. Indeed, during encoding, subsequent remembering was related to stronger ITPC in the hippocampus in our EEG data, which is consistent with the model-simulated data. Alternatively, the effect might be induced by multisensory integration in primary sensory or higher-level association regions such as superior temporal sulcus or PFC,[37] which then have a knock-on effect on regions downstream. This would be consistent with findings showing STDP to be ubiquitous in multisensory or large-scale cortical regions.[38–40] However, the EEG source analysis did not reveal any significant subsequent memory effect in the PFC in our empirical data. Future experiments with good temporal and spatial resolution such as iEEG or MEG may reveal the underlying network of the STDP-like memory effect. Furthermore, future experiments could pharmacologically manipulate the level of N-methyl-D-aspartate (NMDA) receptor activation (e.g., Weise et al.[41]) because STDP is NMDA dependent.[21,26] If the memory effect is attributed to STDP, the difference between the memory performance in the shortest and longest delay conditions should be decreased, that is, the curves should be flattened when NMDA blockers are applied. Equally important, our study offers a precise and practical method to study the behavioral consequences of synaptic changes in human brain, which bridges the gap between the *in vitro* studies and human episodic memory.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
  - Materials
  - Procedure
  - Computational model
- METHOD DETAILS
  - EEG recordings and preprocessing
  - Unimodal source localization
  - Multimodal source reconstruction



(legend on next page)

- QUANTIFICATION AND STATISTICAL ANALYSIS
    - ○ Realigning condition labels
    - ○ Whole brain subsequent memory effect
    - ○ Additional Analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.06.062.

ACKNOWLEDGMENTS

This research was supported by grants from the European Research Council (https://erc.europa.eu/, Consolidator Grant 647954 to S.H.), Wolfson Foundation and Royal Society (https://royalsociety.org/grants-schemes-awards/grants/wolfson-researchmerit/; to S.H.), and the Economic Social Sciences Research Council (https://esrc.ukri.org/, ES/R010072/1 to S.H. and K.L.S.). The authors would like to thank George Parish for providing the software code for the computational model, Andrew Clouter for providing the scripts for study materials and experimental programs, and the Cognition and Oscillations Lab at the University of Glasgow and the University of Birmingham for their helpful inputs. The first author would also like to thank Martin I. Antov, Philipp Bierwirth, Elena Plog, and Ursula Stockhorst for their helpful discussions.

AUTHOR CONTRIBUTIONS

Conceptualization, S.H., K.L.S., and D.W.; investigation, formal analysis, and writing – original draft, D.W.; writing – review & editing, funding acquisition, and supervision, S.H. and K.L.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 13, 2023
Revised: May 3, 2023
Accepted: June 22, 2023
Published: July 17, 2023; corrected online: August 10, 2023

REFERENCES

1. Tulving, E. (2002). Episodic memory: from mind to brain. Annu. Rev. Psychol. *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114.
2. Martin, S.J., Grimwood, P.D., and Morris, R.G.M. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. Annu. Rev. Neurosci. *23*, 649–711. https://doi.org/10.1146/annurev.neuro.23.1.649.
3. Neves, G., Cooke, S.F., and Bliss, T.V.P. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. Nat. Rev. Neurosci. *9*, 65–75. https://doi.org/10.1038/nrn2303.
4. Bi, G.Q., and Poo, M.M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J. Neurosci. *18*, 10464–10472. https://doi.org/10.1523/JNEUROSCI.18-24-10464.1998.
5. Testa-Silva, G., Verhoog, M.B., Goriounova, N.A., Loebel, A., Hjorth, J., Baayen, J.C., de Kock, C.P., and Mansvelder, H.D. (2010). Human synapses show a wide temporal window for spike-timing-dependent plasticity. Front. Synaptic Neurosci. *2*, 12. https://doi.org/10.3389/fnsyn.2010.00012.
6. Verhoog, M.B., Goriounova, N.A., Obermayer, J., Stroeder, J., Hjorth, J.J.J., Testa-Silva, G., Baayen, J.C., de Kock, C.P.J., Meredith, R.M., and Mansvelder, H.D. (2013). Mechanisms underlying the rules for associative plasticity at adult human neocortical synapses. J. Neurosci. *33*, 17197–17208. https://doi.org/10.1523/JNEUROSCI.3158-13.2013.
7. Stefan, K., Kunesch, E., Cohen, L.G., Benecke, R., and Classen, J. (2000). Induction of plasticity in the human motor cortex by paired associative stimulation. Brain *123*, 572–584. https://doi.org/10.1093/brain/123.3.572.
8. Wolters, A., Sandbrink, F., Schlottmann, A., Kunesch, E., Stefan, K., Cohen, L.G., Benecke, R., and Classen, J. (2003). A temporally asymmetric Hebbian rule governing plasticity in the human motor cortex. J. Neurophysiol. *89*, 2339–2345. https://doi.org/10.1152/jn.00900.2002.
9. Fu, Y.-X., Djupsund, K., Gao, H., Hayden, B., Shen, K., and Dan, Y. (2002). Temporal specificity in the cortical plasticity of visual space representation. Science *296*, 1999–2003. https://doi.org/10.1126/science.1070521.
10. McMahon, D.B.T., and Leopold, D.A. (2012). Stimulus timing-dependent plasticity in high-level vision. Curr. Biol. *22*, 332–337. https://doi.org/10.1016/j.cub.2012.01.003.
11. Romei, V., Chiappini, E., Hibbard, P.B., and Avenanti, A. (2016). Empowering reentrant projections from V5 to V1 boosts sensitivity to motion. Curr. Biol. *26*, 2155–2160. https://doi.org/10.1016/j.cub.2016.06.009.
12. Yao, H., Shen, Y., and Dan, Y. (2004). Intracortical mechanism of stimulus-timing-dependent plasticity in visual cortical orientation tuning. Proc. Natl. Acad. Sci. USA *101*, 5081–5086. https://doi.org/10.1073/pnas.0302510101.
13. Yao, H., and Dan, Y. (2001). Stimulus timing-dependent plasticity in cortical processing of orientation. Neuron *32*, 315–323. https://doi.org/10.1016/S0896-6273(01)00460-3.
14. Hebb, D.O. (2002). The Organization of Behavior: A Neuropsychological Theory (L. Erlbaum Associates).
15. Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science *275*, 213–215. https://doi.org/10.1126/science.275.5297.213.
16. Hanslmayr, S., Axmacher, N., and Inman, C.S. (2019). Modulating human memory via entrainment of brain oscillations. Trends Neurosci. *42*, 485–499. https://doi.org/10.1016/j.tins.2019.04.004.
17. Clouter, A., Shapiro, K.L., and Hanslmayr, S. (2017). Theta phase synchronization is the glue that binds human associative memory. Curr. Biol. *27*, 3143–3148.e6. https://doi.org/10.1016/j.cub.2017.09.001.

Figure 4. Recall accuracy as a function of single-trial phase offset between visual and auditory activity

(A) Phase differences between visual and auditory grand average ERP (N = 48) in each phase bin. Amplitude was normalized. Instantaneous phase differences were averaged between 0.5 and 2.5 s and plotted by the mean resultant vector on a unit circle with count histogram of the phase differences between 0.5 and 2.5 s. The gray shaded time windows were zoomed in to show the phase delays between the two time series. See also Figure S2 for the procedure of the single-trial analysis.

(B) Recall accuracy in each phase bin is normalized by subtracting the mean and dividing by the standard deviation. Error bars represent SE. *$p < 0.05$. N = 23 for both groups. See also Figure S3B for the results of raw recall accuracy.

(C) Source localization of the subsequent memory effect. Left: the strongest ITPC difference between subsequently remembered trials and forgotten trials was localized, MNI coordinates: −40, −31, and −10. The paired-sample t values were interpolated. Right: t values were only plotted for the hippocampal ROIs (AAL atlas).

(D) The ITPC at the left hippocampus was averaged across 0.5 and 2.5 s and plotted as a function of frequency.

(E) Same as (D), but the ITPC was computed for the simulated LFP data of high weights trials and low weights trials. The ITPC was averaged across auditory and visual groups. See also Figure S4 for the TFRs of the ITPC of the empirical and simulated data.

18. Parish, G., Hanslmayr, S., and Bowman, H. (2018). The sync/deSync model: how a synchronized hippocampus and a desynchronized neocortex code memories. J. Neurosci. *38*, 3428–3440. https://doi.org/10.1523/JNEUROSCI.2561-17.2018.

19. Wang, D., Parish, G., Shapiro, K.L., and Hanslmayr, S. (2021). Interaction between theta-phase and spike-timing dependent plasticity simulates theta induced memory effects. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.24.469900.

20. Song, S., Miller, K.D., and Abbott, L.F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat. Neurosci. *3*, 919–926. https://doi.org/10.1038/78829.

21. Caporale, N., and Dan, Y. (2008). Spike timing–dependent plasticity: a Hebbian learning rule. Annu. Rev. Neurosci. *31*, 25–46. https://doi.org/10.1146/annurev.neuro.31.060407.125639.

22. Di Russo, F., Martínez, A., Sereno, M.I., Pitzalis, S., and Hillyard, S.A. (2002). Cortical sources of the early components of the visual evoked potential. Hum. Brain Mapp. *15*, 95–111. https://doi.org/10.1002/hbm.10010.

23. Picton, T.W. (2010). Human Auditory Evoked Potentials (Plural Publishing).

24. Schnapf, J.L., Kraft, T.W., and Baylor, D.A. (1987). Spectral sensitivity of human cone photoreceptors. Nature *325*, 439–441. https://doi.org/10.1038/325439a0.

25. Wang, D., Clouter, A., Chen, Q., Shapiro, K.L., and Hanslmayr, S. (2018). Single-trial phase entrainment of theta oscillations in sensory regions predicts human associative memory performance. J. Neurosci. *38*, 6299–6309. https://doi.org/10.1523/JNEUROSCI.0349-18.2018.

26. Mansvelder, H.D., Verhoog, M.B., and Goriounova, N.A. (2019). Synaptic plasticity in human cortical circuits: cellular mechanisms of learning and memory in the human brain? Curr. Opin. Neurobiol. *54*, 186–193. https://doi.org/10.1016/j.conb.2018.06.013.

27. Cobb, S.R., Buhl, E.H., Halasy, K., Paulsen, O., and Somogyi, P. (1995). Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. Nature *378*, 75–78. https://doi.org/10.1038/378075a0.

28. Huerta, P.T., and Lisman, J.E. (1995). Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro. Neuron *15*, 1053–1063. https://doi.org/10.1016/0896-6273(95)90094-2.

29. Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., and Schroeder, C.E. (2009). The leading sense: supramodal control of neurophysiological context by attention. Neuron *64*, 419–430. https://doi.org/10.1016/j.neuron.2009.10.014.

30. Müller-Dahlhaus, F., Ziemann, U., and Classen, J. (2010). Plasticity resembling spike-timing dependent synaptic plasticity: the evidence in human cortex. Front. Synaptic Neurosci. *2*, 34. https://doi.org/10.3389/fnsyn.2010.00034.

31. Roux, F., Parish, G., Chelvarajah, R., Rollings, D.T., Sawlani, V., Hamer, H., Gollwitzer, S., Kreiselmeyer, G., ter Wal, M.J., Kolibius, L., et al. (2022). Oscillations support short latency co-firing of neurons during human episodic memory formation. eLife *11*, e78109, https://doi.org/10.7554/eLife.78109.

32. Adaikkan, C., Middleton, S.J., Marco, A., Pao, P.-C., Mathys, H., Kim, D.N.-W., Gao, F., Young, J.Z., Suk, H.-J., Boyden, E.S., et al. (2019). Gamma entrainment binds higher-order brain regions and offers neuroprotection. Neuron *102*, 929–943.e8. https://doi.org/10.1016/j.neuron.2019.04.011.

33. Martorell, A.J., Paulson, A.L., Suk, H.-J., Abdurrob, F., Drummond, G.T., Guan, W., Young, J.Z., Kim, D.N.-W., Kritskiy, O., Barker, S.J., et al. (2019). Multi-sensory gamma stimulation ameliorates Alzheimer's-associated pathology and improves cognition. Cell *177*, 256–271.e22. https://doi.org/10.1016/j.cell.2019.02.014.

34. Chan, D., Suk, H.-J., Jackson, B.L., Milman, N.P., Stark, D., Klerman, E.B., Kitchener, E., Fernandez Avalos, V.S., de Weck, G., Banerjee, A., et al. (2022). Gamma frequency sensory stimulation in mild probable Alzheimer's dementia patients: results of feasibility and pilot studies. PLoS One *17*, e0278412, https://doi.org/10.1371/journal.pone.0278412.

35. Schneider, M., Tzanou, A., Uran, C., and Vinck, M. (2023). Cell-type-specific propagation of visual flicker. Cell Rep. *42*, 112492, https://doi.org/10.1016/j.celrep.2023.112492.

36. Soula, M., Martín-Ávila, A., Zhang, Y., Dhingra, A., Nitzan, N., Sadowski, M.J., Gan, W.-B., and Buzsáki, G. (2023). Forty-hertz light stimulation does not entrain native gamma oscillations in Alzheimer's disease model mice. Nat. Neurosci. *26*, 570–578. https://doi.org/10.1038/s41593-023-01270-2.

37. Werner, S., and Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. *30*, 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010.

38. Basura, G.J., Koehler, S.D., and Shore, S.E. (2015). Bimodal stimulus timing-dependent plasticity in primary auditory cortex is altered after noise exposure with and without tinnitus. J. Neurophysiol. *114*, 3064–3075. https://doi.org/10.1152/jn.00319.2015.

39. Casula, E.P., Pellicciari, M.C., Picazio, S., Caltagirone, C., and Koch, G. (2016). Spike-timing-dependent plasticity in the human dorso-lateral prefrontal cortex. NeuroImage *143*, 204–213. https://doi.org/10.1016/j.neuroimage.2016.08.060.

40. Marks, K.L., Martel, D.T., Wu, C., Basura, G.J., Roberts, L.E., Schvartz-Leyzac, K.C., and Shore, S.E. (2018). Auditory-somatosensory bimodal stimulation desynchronizes brain circuitry to reduce tinnitus in guinea pigs and humans. Sci. Transl. Med. *10*, eaal3175, https://doi.org/10.1126/scitranslmed.aal3175.

41. Weise, D., Mann, J., Rumpf, J.-J., Hallermann, S., and Classen, J. (2016). Differential regulation of human paired associative stimulation-induced and theta-burst stimulation-induced plasticity by L-type and T-type Ca^{2+} channels. Cereb. Cortex *27*, 4010–4021. https://doi.org/10.1093/cercor/bhw212.

42. Chen, Q., Wang, D., Shapiro, K.L., and Hanslmayr, S. (2021). Using fast visual rhythmic stimulation to control inter-hemispheric phase offsets in visual areas. Neuropsychologia *157*, 107863, https://doi.org/10.1016/j.neuropsychologia.2021.107863.

43. Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436. https://doi.org/10.1163/156856897X00357.

44. Kleiner, M., Brainard, D., and Pelli, D.G. (2007). What's new in psychtoolbox-3? Perception *36*, 1–16.

45. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. *10*, 437–442. https://doi.org/10.1163/156856897X00366.

46. Jensen, O., Idiart, M.A., and Lisman, J.E. (1996). Physiologically realistic formation of autoassociative memory in networks with theta/gamma oscillations: role of fast NMDA channels. Learn. Mem. *3*, 243–256.

47. Graupner, M., and Brunel, N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. Proc. Natl. Acad. Sci. USA *109*, 3991–3996. https://doi.org/10.1073/pnas.1109359109.

48. O'Reilly, R.C., Bhattacharyya, R., Howard, M.D., and Ketz, N. (2014). Complementary learning systems. Cogn. Sci. *38*, 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x.

49. Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., and Leahy, R.M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. *2011*, 879716, https://doi.org/10.1155/2011/879716.

50. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869, https://doi.org/10.1155/2011/156869.

51. Huiskamp, G. (1991). Difference formulas for the surface Laplacian on a triangulated surface. J. Comput. Phys. *95*, 477–496. https://doi.org/10.1016/0021-9991(91)90286-T.

52. Oostendorp, T.F., and van Oosterom, A. (1996). The surface Laplacian of the potential: theory and application. IEEE Trans. Bio Med. Eng. *43*, 394–405. https://doi.org/10.1109/10.486259.

53. Murzin, V., Fuchs, A., and Scott Kelso, J.A. (2013). Detection of correlated sources in EEG using combination of beamforming and surface Laplacian methods. J. Neurosci. Methods *218*, 96–102. https://doi.org/10.1016/j.jneumeth.2013.05.001.

54. Van Veen, B.D., Van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Bio Med. Eng. *44*, 867–880. https://doi.org/10.1109/10.623056.

55. Berens, P. (2009). CircStat: a M*ATLAB* toolbox for circular statistics. J. Stat. Softw. *31*, 1–21. https://doi.org/10.18637/jss.v031.i10.

56. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

57. Brown, V.A. (2021). An introduction to linear mixed-effects modeling in R. Psychol. Sci. *4*, 1, https://doi.org/10.1177/2515245920960351.

58. Barr, D.J. (2021). Learning statistical models through simulation in R: an interactive textbook. https://psyteachr.github.io/book/ug3-stats.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MATLAB	https://www.mathworks.com	R2020a
Psychophysics Toolbox	http://psychtoolbox.org	3
ActiView	http://www.biosemi.com	7
FieldTrip	https://www.fieldtriptoolbox.org	20201229
Audacity	https://www.audacityteam.org	2.1.2
MPEG Streamclip	http://www.squared5.com	1.0.3b8 beta for Mac OS X
Brainstorm	https://neuroimage.usc.edu/brainstorm	N/A
Sync-deSync-model-TIME-STDP	This paper	https://osf.io/fpyqk/
RStudio	https://posit.co/products/open-source/rstudio/	2023.03.1+446
lme4	https://cran.r-project.org/web/packages/lme4/index.html	1.1-32
Circular Statistics Toolbox	https://www.mathworks.com/matlabcentral/fileexchange/10676-circular-statistics-toolbox-directional-statistics	1.21.0.0
Other		
Iiyama Vision Master Pro514 HM204DT	https://iiyama.com	CRT monitor
nVidia Quadro K620	https://www.nvidia.com	Graphics card
ER3C	https://www.etymotic.com	Insert earphones
ThorLabs DET36A	https://www.thorlabs.com	Photodiode
BioSemi ActiveTwo system with Analog Input Box	http://www.biosemi.com	EEG system
Fastrak	https://polhemus.com	Electromagnetic digitizer

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Simon Hanslmayr (simon.hanslmayr@glasgow.ac.uk) danying.wang@glasgow.ac.uk.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The datasets are available from: https://osf.io/fpyqk/.
- All original code has been deposited at https://osf.io/fpyqk/. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

In total, 107 healthy English-speaking young adults participated in the experiments. In the auditory cue group, 51 participants (35 females; mean age: 19.6 years; range: 18 – 32 years) performed the experiment. Six participants were left-handed. The remaining

45 participants were right-handed. 34 participants were given course credits via the University of Birmingham's Psychology Research Participation Scheme. The remaining participants were paid £8 per hour for their participation. The data from three participants were excluded because of chance-level memory performance. The data from the remaining 48 participants were retained for the final data analysis. In the visual cue group, 55 participants (35 females; mean age: 25 years; range: 18 – 40 years) participated in the experiment. 51 participants had not participated in the auditory cue group. Two participants who participated in the auditory cue group took part in the experiment as the study was originally designed as a within-group design. However, because of the COVID-19 pandemic, participants were not able to return. Another two participants participated in one of the pilot experiments of the auditory cue group and their data was only included for the visual cue group. Four participants were left-handed. One participant was ambidextrous. 50 participants were right-handed. Apart from seven participants who were granted course credits, the remaining 48 participants were paid £8 per hour for their participation. The data from seven participants were excluded due to chance-level memory performance. The data from the remaining 48 participants were retained for the final EEG data analysis. The EEG data from one participant were excluded due to poor EEG data quality (less than 15 trials were survived after artefacts rejections per condition). The behavioral data analysis included 46 participants who only participated in each group once if a between subject factor was included in the analysis. All participants had normal or corrected-to-normal vision and normal hearing.

Materials

The visual stimuli were taken from the same stimulus set as those used in Clouter et al.,[17] Wang et al.,[25] and Chen et al.[42] Some of the auditory stimuli were from the same set as those used in Clouter et al.[17] and Wang et al.[25] The remainder of the auditory stimuli were acquired from additional tracks of Apple Loops for Garage Band and two unique soundtracks that were royalty free. Movie clips of 3 s each had 227 frames in total with a frame frate of 75 frames/s, which was converted from the original 25 frames/s using MPEG Streamclip (http://www.squared5.com/). The movie clips were taken from documentaries depicting natural scenes, animals, architectures or human activities. 288 movie clips were modulated at 37.5 Hz with luminance changing between 0% and 100% (initially starting at 100% luminance). All 288 sound clips were preprocessed using Audacity software (2.1.2 https://www.audacityteam.org/) as in Clouter et al.[17] Each sound clip was presented concurrently with the presentation of a movie for 3 s, with a lag of 40 ms, which compensated for the fact that auditory stimuli are processed faster than visual stimuli Clouter et al.[17] Sound amplitude was modulated at 37.5 Hz from 0 to 100% with a sine wave, at 0°, 90°, 180° and 270° phase offsets from the movies. The movie clips and sound clips were also modulated at 4 Hz. Each sound was modulated at 0° and 180° phase offsets from the 4 Hz sine-wave-modulated movies. The sound clips were not directly related to the contents of the videos. Each sound clip was taken from one of the eight sound categories as described in Wang et al.[25] Each sound category had 36 sound clips. All sounds were randomly divided into six sets of equal size (48 sounds per set), with the constraint that the number of sounds for each sound category was equal. The assignment of presentation frequency (37.5 Hz: 4 sets 192 sounds or 4 Hz: 2 sets 96 sounds) and phase offset conditions to each sound set were counterbalanced across participants. For each participant, a movie was assigned randomly at 37.5 Hz (192 movies), or 4 Hz (96 movies), then randomly assigned to a sound that was chosen to form a sound-movie presentation pair.

The experimental apparatus and stimulus presentation were identical to that used by Clouter et al.[17] and Wang et al.[25] The experiment was programmed with MATLAB (MathWorks) using the Psychophysics Toolbox extensions.[43–45] Presentation of visual stimuli were on a 21-inch CRT display (Iiyama Vision Master Pro514 HM204DT) with an nVidia Quadro K620 graphics card (1058 MHz graphics clock, 2048 MB dedicated graphics memory, NVidia). Participants sat ~60 cm from the center of the monitor. 53 participants recorded with EEG while resting their head on a chin rest. The monitor screen refresh rate was 75 Hz. Auditory stimuli were presented with insert earphones (ER-3C, Etymotic Research). The physical presentation of phase offsets between movies and sounds, as well as the frequencies of the movies and sounds, were verified by a ThorLabs DET36A photodiode (https://www.thorlabs.com/) and a line-out speaker using 3.5 mm audio connectors connecting with a Biosemi Analogue Input Box (https://www.biosemi.com/aib.htm).

Procedure

Participants in both the visual and auditory cue groups provided informed consent and were given task instructions, before practicing the procedure with four example trials. In the visual cue group, 29 participants were prepared for EEG data collection. The remainder of the participants in the visual cue group performed the behavioral tasks without EEG being recorded. In the auditory cue group, 24 participants' EEG were recorded. 27 participants did the behavioral tasks without EEG recording. During the formal experiment, participants were monitored by a web camera connected to a monitor in the control room. Participants were asked to wave at the web camera if they had any questions or requests during the experiment.

The visual cue group were presented with 16 blocks of an associative memory task, where stimuli were modulated at 37.5 Hz. Each associative memory task block consisted of an encoding phase, a distractor phase, and an associative memory recall test phase. During the encoding phase, the procedure was the same as described in Wang et al.[25] Participants were presented with a movie along with a sound for each trial. Each trial started with a fixation cross, which served as inter-trial-interval and lasted between 1 and 3 s. Then, the sound-movie pair was presented for 3 s. Participants were instructed to press one of the five number keys on a keyboard, to indicate how well the sound suited the contents of the movie after the presentation of the sound-movie pair. The instruction screen was presented until a response was made. The ratings ranged from 1 (the sound does not suit the contents of the movie at all) to 5 (the sound suits the movie very well). Participants were instructed to remember the association between the sound and the movie. Each block consisted of 12 trials. Four sounds from three categories were associated with the 12 movies with the constraint that the number of sounds for each phase offset condition was equal (i.e., one for 0°, one for 90°, one for 180° and one for 270° in one sound category). For participants whose EEG was recorded, another 8 blocks followed with only the encoding phase, during which the stimuli were modulated at 4 Hz. Participants were instructed to make a judgment as to how well the sound suited the contents of the movie but no memory test later on. Those blocks served as ground truth for the analysis of phase offsets between auditory and visual sources after adjusting for dipole orientation (see method details section multimodal source reconstruction).

The distractor phase was the same for both cue groups, as was done by Clouter et al.[17] and Wang et al.[25] During this phase, participants were presented with a random number that was drawn from 170 to 199 and instructed to count aloud backward from this number in steps of 3 for 30 seconds.

The associative memory recall test commenced following the distractor phase. In the visual cue group, the test phase consisted of 12 trials. Each trial started with a fixation cross between 1 and 3 seconds. Participants were presented with one of the 12 movies presented during the encoding phase for 3 s and instructed to recall the paired sound. Then, participants were presented with four sounds from the encoding phase and free to choose the order of which sound they would like to hear, using the number keys 1 through 4. After they completed listening to all four options, they were instructed to select the sound that they thought was played with the movie in the encoding phase, using the number keys 1 through 4. In both stages, the screen was presented until a response was made. The sounds from which to choose were all from the same sound category.

The associative memory task in the auditory cue group was similar to that of the visual cue group, except that each task block consisted of 16 trials and in total there were 12 associative memory task blocks. The stimuli in the 12 blocks were modulated at 37.5 Hz. The following six blocks consisted of the encoding phase only and the stimuli were modulated at 4 Hz if participants' EEG was recorded. The different trial numbers between different cue groups were implemented because pilot results showed these trial numbers enabled participants to achieve acceptable memory performance and shortened the time required to complete the experiment. The encoding phase in each block of the auditory cue group was the same as in each block of the visual cue group. The associative memory recall test phase was identical to that employed by Wang et al.[25] Participants were tested for all 16 trials within each block. Each trial started with a fixation cross randomly chosen to appear for between 1 and 3 s. Participants were presented with 1 of the 16 sounds presented during the encoding phase for 3 s, along with four still images from the four movies presented during the encoding phase. Participants were instructed to select the paired movie using the number keys 1 through 4. The instruction screen was presented until a response was made. The movies from which to choose were, in the encoding phase, all presented with a sound from the same sound category.

In both cue groups, two unimodal source localizer tasks were conducted following the multimodal blocks for those participants for whom EEG was recorded. The unimodal source localizer tasks served to separate sources from each sensory modality for multimodal source reconstruction. The tasks were exactly same as those used by Wang et al.[25] The unimodal auditory task consisted of 50 trials of 4 Hz modulated sound clips. The unimodal visual task consisted of 50 trials of 4 Hz modulated movie clips. The unimodal stimuli of both modalities were presented for 3s. Participants were asked to rate how pleasant each sound or movie was using the number keys 1 (the sound or the movie was very unpleasant) through 5 (the sound or the movie was very pleasant) for each trial. The response to the 37.5 Hz stimulus was expected to originate from the same sensory source as the response from the 4 Hz stimulus. Only the unimodal localizers from the 4Hz modulated stimuli were used since we have piloted with 37.5 Hz and 4 Hz unimodal modulation and the results suggested that signal-to-noise ratio for 4 Hz steady-state evoked responses was higher than for 37.5 Hz responses.

For all participants, 24 sound-movie pairs randomly drawn from the associative memory task blocks were presented to test participants' perception about the synchrony between a sound and a movie as the last task of the experiments. The stimuli were modulated at 37.5 Hz. Participants were asked whether they could detect if the change of the modulated auditory stimulus was in synchrony with the corresponding modulated luminance of a movie (0° phase offset) or out-of-synchrony (90°, 180° or 270° phase offsets). Participants were instructed to press the number keys 1 for out-of-synchrony and 2 for in synchrony.

Computational model

We adapted a computational neural network model from Wang et al.,[19] which comprises two groups of neurons that represent the neo-cortex (NC) and the hippocampus. Each group of neurons was split into two subgroups that represent the visual and auditory stimulus, respectively (N_{nc}= 20, N_{hipp}=10). The neuron physiology is simulated as in Wang et al.[19] except that only the STDP learning rule was retained. There was no hippocampal theta learning system. Specifically, neuron membrane potential changes are simulated using an integrate-and-fire equation (Equation 1), where the membrane potential decays over time to a resting potential (E_L = -70mV) at a rate dictated by the membrane conductance (g_m = 0.03). Here, a spike event is generated if the voltage exceeds a threshold (V_{th} = -55mV), at which time the voltage is clamped to the resting potential for an absolute length of time to approximate a refractory period (2ms). As well as the leak current, the input current for model neurons contains the sum of all spike events occurring at pre-synaptic neurons (I_{syn}), alternating current (AC) that represents NC alpha oscillations (I_{AC}), any existing direct current (I_{DC}) and an after-depolarization (ADP) function (I_{ADP}), described subsequently.

$$C_m \frac{dV_m}{dt} = g_m(E_L - V_m) + I_{syn} + I_{AC} + I_{DC} + I_{ADP} \qquad \text{(Equation 1)}$$

The leaky integrate and fire equation.

Equation 2 explains the process by which neurons communicate through spike events, whereby the sum of all spike events over time makes up the I_{syn} current. Here, an alpha function is used to model the excitatory post-synaptic potential (EPSP), which provides an additive exponential function that diminishes the further the current time point (*t*) is from the initiating spike event (t_{fire}). The amplitude of the function is dictated by the current synaptic weight of the post-synaptic synapse ($0 \leq \rho \leq 1$) multiplied by its maximal weight (W_{max}). All spike events had a delay of 2 ms before they reached post-synaptic connections.

$$EPSP(t) = W_{max} \cdot \rho(t) \cdot \left(e \cdot \frac{\Delta t}{\tau_s} \right) \cdot e^{-\frac{\Delta t}{\tau_s}}, \Delta t = t - t_{fire} \qquad \text{(Equation 2)}$$

Generation of an excitatory post-synaptic potential (EPSP) through time using an alpha-function.

Hippocampal neurons received additional input from an ADP function, as in previous models[18,46]; Equation 3; A_{ADP} = 0.2nA, τ_{ADP} = 250ms). This provided exponentially ramping input, which was reset after each spike-event (t_{fire}).

$$I_{ADP}(t) = \frac{A_{ADP} \cdot \Delta t}{\tau_{ADP}} \cdot e^{1 - \frac{\Delta t}{\tau_{ADP}}}, \Delta t = t - t_{fire} \qquad \text{(Equation 3)}$$

After-depolarization (ADP) function.

The learning rule was implemented via an adapted spike-time-dependent plasticity (STDP) mechanism, inspired by other models.[20,47] We consider two bi-directionally connected neurons in a traditional STDP framework. Upon the occurrence of a spike event in a model neuron, post-synaptic weights are strengthened for any given pre-synaptic neuron that spiked beforehand or weakened in the vice versa condition; the assumption being that the spike arriving at the post-synaptic connection must have either contributed to or competed with the spike event in question, depending on the directionality of the connection, leading to a reward or punishment of the synapse, respectively. To implement this, we calculate potential synaptic plasticity via functions for long-term potentiation (F_{LTP}) and long-term depression (F_{LTD}) at the time of an eliciting spike (t) in Equations 4 and 5.

In the case of potentiation (Equation 4), potential LTP at the post-synaptic connection (i) is calculated as the summation of historic pre-synaptic spikes (n_{pre}) that occurred before the spike event in question (where $t_i < t$), weighted by an absolute value ($A_+ = 0.65$). Contributions of pre-synaptic spikes were proportional to an exponential decay, thus favouring spikes that occurred close together in time (τ_S = 20ms). In the case of depression (Equation 5), potential LTD at the pre-synaptic connection (j) was similarly calculated as the summation of historic post-synaptic spikes (n_{post}) that occurred before the spike event in question (where $t_j < t$), weighted by an absolute value ($A_- = 0.65$).

$$F_{LTP}(t,i) = \sum_{t_i<t}^{n_{pre}} A_+ \cdot e^{\frac{t_i - t}{\tau_S}} \qquad \text{(Equation 4)}$$

$$F_{LTD}(t,j) = \sum_{t_j<t}^{n_{post}} A_- \cdot e^{\frac{t_j - t}{\tau_S}} \qquad \text{(Equation 5)}$$

Synaptic plasticity functions (F) calculate potential plasticity as the summation of the total number (n_{post} & n_{pre}) of historic spike events (t_i & t_j) arriving at a post- (i) or pre-synaptic (j) synapse relative to a given spike event (t), where an absolute value (A_+/A_-) is modulated by the difference in spike times and theta phase

Neurons within each subgroup (i.e., auditory, or visual) of the NC had a 25% chance of being connected (W_{max} = 0.3). Connections of neurons between subgroups were not implemented in NC as it was assumed visual and auditory stimuli had not been previously associated. Synaptic plasticity was also considered not to be operating on cortical synapses as in the complimentary systems framework[48] it is assumed that cortical plasticity occurs on a much slower timescale. Background noise for each NC neuron was estimated by Poisson distributed spike-trains (4000 spikes/s, W_{max} = 0.023). A cosine wave of 10 Hz (amplitude = 0.1pA) was fed into NC neurons via I_{AC}. Two constant inputs were fed into each NC subgroup to simulate presentation of visual and auditory stimuli via I_{DC} (amplitude = 1.75pA). These inputs were modulated by a cosine wave at 37.5 Hz with four phase offsets.

The two subgroups of hippocampal neurons that represented visual and auditory stimuli were fully connected to their NC counterparts (W_{max} = 0.35 for NC → Hip & W_{max} = 0.08 for Hip → NC synapses), as it was assumed both stimuli were previously known. Background noise for each hippocampal neuron was estimated by Poisson distributed spike-trains (1500 spikes/s, W_{max} = 0.015). Synapses within the entire hippocampus had a probability of 50% of forming a connection (W_{max} = 0.65), such that weights for intra-subgroup synapses were set to maximum and those for inter-subgroup synapses were initially set to 0. Synaptic plasticity was in effect on all hippocampal synapses, allowing for the association of visual and auditory stimuli to take place within the hippocampus.

Two cosine waves ($-1 \leq$ amplitude ≤ 1pA) were fed into the visual and auditory NC subgroups. The stimulus presentation length was three seconds (3000 data points). A 2-second inter-stimulus interval was used before visual-auditory stimulus presentation. The two cosine waves were modulated at 37.5 Hz with auditory stimulus phase offsets of 0°, 90°, 180°, and 270° from the visual stimulus (stimulus strength = 3). Pink noise was added to the two cosine waves. The simulations were run for 192 trials for each condition, which were then averaged across trials for each condition. For each simulation, we randomized a new set of initial synaptic connections as well as new Poisson distributed spike-trains for all conditions. To evaluate the recall performance of the model, the hippocampal weights after learning were averaged between 2.75 to 3 seconds after stimulus onset.

The local field potential (LFP) was computed as in Parish et al.[18] which first measures the activity of a group of neurons aggregating spikes through time. Then it was filtered by a Hanning filter with a 25 ms window. The high weights trials and low weights trials were categorized by pooling all trials across conditions and median split according to the weight changes from the auditory group to the visual group and from the visual group to the auditory group. The ITPC was computed using the same parameters as done in the EEG analysis for the LFP in the auditory group in trials of high and low weight change from visual to auditory group, as well as the LFP in the visual group in trials of high and low weight change from auditory to visual group (see Figure S4 for the Time Frequency Representations of the ITPC in each group). To compare the results with the empirical data from the EEG, where a separation between auditory and visual connections is not possible, the ITPC was averaged between the auditory and visual groups.

METHOD DETAILS

EEG recordings and preprocessing

For participants whose EEG was recorded, 128 scalp channels of a BioSemi ActiveTwo system were used. Vertical eye movements were recorded from an additional electrode placed 1 cm below the left eye. Horizontal eye movements were recorded from two additional electrodes placed 1 cm to the left of the left eye and to the right of the right eye. Analogue signals of a photodiode that was attached to a white square informing the onset of a movie and two audio output channels of a sound were recorded by the BioSemi Analog Input Box, which resulted in three analogue signal channels (one for visual stimuli, two for auditory stimuli). These signals were recorded to later correct for the onset of an EEG trigger, thus redefining the onset of an epoch by its actual onset. This is crucial for steady-state responses in higher frequency as a jitter of onset times might cause inaccurate phase and amplitude information of the fast evoked responses. Online signals were sampled at 2048 Hz using the BioSemi ActiView software. The position of each participant's electrodes was tracked using a Polhemus FASTRAK device (Colchester) and recorded by Brainstorm[49] implemented in MATLAB.

EEG data were preprocessed with the Fieldtrip toolbox.[50] The raw data less than 5 Gb were bandpass filtered between 1 and 120 Hz and bandstop filtered between 48-52 and 98-102 Hz to remove potential line noise at 50 and 100 Hz, then epoched from 2000 ms before stimulus onset to 5000 ms after stimulus onset. The raw data that were larger than 5 Gb were epoched first, then bandpass and bandstop filtered. The epoched data were downsampled to 512 Hz. An independent component analysis (ICA) was implemented after coarse artefact rejection of bad channels and trials by visual inspection. In the multimodal condition of the visual cue group, bad channels were excluded in five participants, with the average number of excluded channels being 4.2 (range 1 to 9). In the unimodal condition, bad channels were excluded in two participants (mean, 2.5, range 2 to 3). In the multimodal condition of the auditory cue group, bad channels were excluded in 10 participants (mean, 2.5, range 1 to 7). In the unimodal condition, bad channels were excluded in six participants (mean, 3.3, range 1 to 9). ICA components that indicated eye blinks and horizontal eye movements and regular pulse artefacts were removed from the data. The bad channels were interpolated by the method of triangulation of nearest neighbors based on the individuals' electrode positions. After average re-referencing, trials that had artefacts were manually rejected by visual inspection. Participants with less than 15 artefact-free trials in any of the conditions of interest were excluded from further analysis. The mean numbers of trials remaining in each condition of the visual cue group and auditory cue group were listed as a Table below (values inside of brackets indicate range):

	Visual cue group	Auditory cue group
Unimodal sound	38 (22-49)	38 (21-48)
Unimodal movie	38 (23-50)	38 (21-48)
Multimodal 37.5 Hz 0°	35 (22-46)	37 (26-45)
Multimodal 37.5 Hz 90°	34 (22-44)	38 (27-46)
Multimodal 37.5 Hz 180°	36 (19-46)	37 (21-44)
Multimodal 37.5 Hz 270°	35 (23-44)	38 (20-46)
Multimodal 4 Hz 0°	38 (24-47)	39 (28-46)
Multimodal 4 Hz 180°	38 (19-47)	38 (20-47)

Since no individual MRI T1 structural scans were available, individuals' electrode positions were aligned to a template head model. Then, source models were prepared with a template volume conduction model and the aligned individuals' electrode positions.

Unimodal source localization

The unimodal source localization was implemented in the same manner as described by Wang et al.[25] The EEG data in the unimodal sound condition were transformed to scalp current density (SCD) using the finite-difference method.[51,52] The leadfields were also SCD transformed by applying the transformation matrix that was used for the SCD transformation.[53] Source activity was reconstructed using a linearly constrained minimum variance (LCMV) beamforming method.[54] Time series SCD data were reconstructed in 2020 virtual electrodes for each participant. Source analysis was conducted with the SCD-transformed leadfields on the SCD-transformed data. In the unimodal movie condition, source analysis was implemented with leadfields that were computed based on scalp potentials. Time series potentials data were reconstructed in virtual electrodes for each participant. Event-related potential (ERP) was calculated at each virtual electrode for each unimodal condition. Time-frequency analysis was applied to the ERPs with a Morlet wavelet (width=7). Evoked power was averaged between 0.75 and 2.75 s after stimulus onset and between 3.5 and 4.5 Hz. A baseline condition was generated by randomly assigning each trial to 0°, 90°, 180° or 270° phase offset by moving the signal onset forward in time by 0, 32, 64 or 96 samples, which correspond to 0, 62.5, 125 or 187.5 ms with the constraint that the numbers of trials in each phase offset were approximately equal. The evoked power in the baseline condition was averaged between 0.75 and 2.75 s and between 3.5 and 4.5 Hz. The evoked power in each unimodal condition was normalized by subtracting the baseline evoked power from the condition evoked power and then divided by the baseline evoked power. This normalized evoked power was grand averaged across 48 participants of both visual cue and auditory cue groups. The grand average evoked power was interpolated to the MNI MRI template. The coordinates for the auditory and visual ROIs were determined by the locations of the maximum grand average evoked power.

Multimodal source reconstruction

The multimodal source reconstruction was implemented as described by Wang et al.[25] except for the steps of flipping the sign (i.e. adjusting the orientation of the dipoles) of reconstructed time series data. The multimodal data were SCD-transformed to reconstruct the time series data from auditory source to get cleaner time series source data from beamforming highly correlated sources.[25,53] Two sets of spatial filters based on the scalp electrodes over the right and left hemisphere, respectively, were computed. The SCD-transformed time series data was applied to the two sets of spatial filters and was extracted at the left auditory ROI and right auditory ROI which were predefined from the unimodal source localization results. The time series data at the visual ROI was reconstructed without SCD transformation and was extracted at the visual ROI.

To solve the sign ambiguity (i.e. dipole orientation) problem caused by beamforming source reconstruction, the signs of reconstructed time series from each source were manually adjusted. The steps were as follows: first, the spatial filters that were extracted from three sensory ROIs were plotted on scalp. The spatial weights showed a preferred scalp distribution of corresponding sensory stimulation. Then, the signs of time series from the ROIs were set to be same as where the largest weights were distributed on the scalp, e.g., the time series data from the visual source would be applied by -1 if the occipital areas where the spatial weights were maximum showed a distribution of negative values. The sign would not be changed if the values of where the spatial weights were maximum were positive. The procedure was applied to all time series data regardless of experimental condition. The time series data from the auditory sources of the two hemispheres were averaged. A sanity check was performed using the data of the 4 Hz modulated conditions. The results were consistent with findings from Clouter et al.[17] and Wang et al.[25] in showing that the instantaneous phase difference in the 0° phase offset condition showed a preferred direction of 0° whereas the 180° condition showed a preferred direction of 180°. A control analysis was also performed to the data of 37.5 Hz modulated conditions to show that the phase relationships between 0° phase offset condition and other phase offset conditions in the auditory ROI were concentrated at 90°, 180°, and 270°. The phase differences between 0° phase offset condition and other phase offset conditions in the visual ROI always showed preferences at 0°. These results were consistent as the physical modulation of the sensory stimuli and remained constant between before and after the sign-flipping procedure. Therefore, the flipping procedure did not bias the results in the direction of our hypothesis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Realigning condition labels

To compute the instantaneous phase differences, the source reconstructed time series data were bandpass filtered between 35 and 40 Hz. The stimulus onset (0 time point) was then redefined by the actual stimulus onset timing measured by the photodiode. The redefined epochs were cut by 1 s before stimulus onset and 4 s after stimulus onset. The ERPs were computed for each phase offset condition at each source. The Hilbert transformation was applied to the source grand averaged ERPs. The instantaneous phase differences were calculated between the unwrapped instantaneous phases from visual and auditory sources for 2 s, beginning 0.5 s after stimulus onset and ending 2.5 s after stimulus onset to avoid influences of onset and offset responses at gamma frequency. Rayleigh and V tests were used to test circular uniformity of the instantaneous phase differences in each phase offset condition. The condition labels were then realigned according to the actual instantaneous phase differences that each condition showed. The circular plots and statistics were done using the Circular Statistics Toolbox for Matlab.[55] The ITPC analysis was done using the Fieldtrip toolbox. Time-frequency analysis was applied for each epoch in the realigned phase conditions at each source using multitaper time-frequency transformation based on Slepian sequences as tapers. The width of smoothing frequency was 1 Hz. The length of a sliding time-window was 1 s. The complex Fourier-spectra was computed using these parameters for frequencies of interest between 20 and 55 Hz and time of interest between -1 and 4 s. The complex Fourier-spectra was then normalized by its magnitude. The ITPC was computed by taking the magnitude of the mean of the normalized complex Fourier-spectra across trials. The ITPC for each condition at each source was grand averaged across 48 participants in both the visual cue and auditory cue groups. To compare the ITPC in each condition between each cue group at each source, the ITPC for each participant was averaged between 0.5 and 2.5 s and between 37 and 38 Hz. To investigate the correlation between the ITPC and recall accuracy across participants, the difference of the mean ITPC between realigned 90° and 270° conditions in which the visual stimulus led in the visual cortex was taken for the visual cue group. The difference of the mean ITPC between realigned 90° and 270° conditions in which the auditory stimulus led in the auditory cortex was taken for the auditory cue group. Similarly, the difference of recall accuracy between realigned 90° and 270° conditions in which the visual stimulus led was taken for the visual cue group. The difference of recall accuracy between realigned 90° and 270° conditions in which the auditory stimulus led was taken for the auditory cue group. The Pearson correlation coefficient was calculated between the ITPC difference and the recall accuracy difference across 48 participants regardless of cue groups. The numbers of trials (see the table above) in each condition contributed to the ITPC analysis did not differ statistically significant, F(3, 141) = 1.196, p = 0.314. The mean difference in trial numbers between realigned 90 and 270 conditions was -1 (range from -8 to 7), t(47) = -1.715, p = 0.093 (two-tailed). The Pearson correlation coefficient was calculated between the trial number difference and the recall accuracy difference, r = 0.022, p = 0.83, two-tailed, suggesting neither the difference between conditions in ITPC nor the correlation was due to the difference in trial numbers.

On the single trial level, the instantaneous phase differences were calculated between the instantaneous phases from visual and auditory sources for the same length. The mean angle direction was computed across these data points. The angle values were sorted from -pi to pi and then divided into four equal sized bins (i.e. each bin containing the exact same amount of trials). If there were remainders after dividing the trial numbers by four, the same number of the last few trials whose angle values were closest to pi were discarded from the bin that had the largest angle values. Mean trial number in each bin for 48 participants is 36 (range from 23 to 45). Mean of the remainder trials is 1.6 (range from 0 to 3). The proportion of remembered trials in each phase bin was calculated. The proportion of remembered trials was normalized by subtracting the mean across all phase bins and divided by the standard deviation (STD). For each participant, trials were randomly drawn to be assigned to four equal sized bins. STD was calculated across the four bins. The procedure was repeated for 100 times and a mean STD was used for the normalization. This procedure effectively scaled the differences between phase bins based on variability of the data, thus down-scaling effects of subjects with fewer trial numbers. A generalized linear mixed effects model was built using RStudio and the package lme4[56] following the tutorial by Brown[57] and Barr.[58] The full model included the interaction between ‘phase bin’ and ‘cue condition’ and all lower-order terms including a random effect ‘subjects’ and control parameters of the Nelder Mead optimizer and removal of the derivative calculations to resolve the non-convergence of the model. The reduced model included all terms in the full model except the interaction. The model comparison was done using the likelihood-ratio test. The model included 48 participants (24 for each group including one participant who participated in both groups) and confirmed a better fit for the full model, $\chi^2(1) = 6.28$, $p = 0.012$. The ERPs were computed for each phase bin at each source. To check whether each phase bin showed preferred phase angle directions at 0°, 90°, 180° and 270°, the same procedure for computing the instantaneous phase differences between grand averaged ERPs at each source as described above was applied.

Whole brain subsequent memory effect

The multimodal data were SCD-transformed to localize possible correlated sources such as the left and right hippocampi (AAL for SPM8 Version V1[53]). The ITPC was calculated separately for subsequently remembered and forgotten trials during the encoding phase, and averaged between 0.5 and 2.5 s and between 37 and 38 Hz. The numbers of remembered and forgotten trials did not differ statistically significant, t(47) = 0.19, p = 0.847 (two-tailed, mean of remembered trials: 73 (range from 36 to 133); mean of forgotten trials: 72 (range from 39 to 107)). Therefore, the difference in the ITPC between remembered and forgotten trials was not due to the difference between the trial numbers. The paired-sample t statistics was interpolated to the template MRI. Then the statistical comparisons between the grand averaged ITPC (N = 48) of the remembered trials and forgotten trials were conducted by masking the ROIs including the left and right hippocampi using a two-tailed paired-samples permutation test with the Monte Carlo method and 1000 randomizations. The same analysis was applied to the ROIs including the left and right superior frontal gyri, the left and right middle frontal gyri and the left and right inferior frontal gyri (atlas defined by AAL for SPM8 Version V1[53]). The significance level was set to 0.05.

Additional Analyses

We analyzed participants’ subjective rating on how well a given sound suited the content of the corresponding video. A main effect of subsequent memory revealed that the mean rating for subsequently remembered pairs was significantly higher than for subsequently forgotten pairs, $F(1, 94) = 47.113$, $p < 0.001$, reflecting that video-sound pairs that were perceived as congruent were remembered better. When relabeled our experimental conditions based on the actual phase offsets in each experimental condition, the rating did not differ between the visually cued group and the auditorily cued group for the 90° phase offset condition, which suggests the memory effect observed in the 90° phase offset condition was not driven by higher semantic congruency between video-sound pairs in the visually cued recall group. When sorting the trials into the phase bins based on single trial phase offset values between visual and auditory activity, the analysis of participants’ subjective rating on how well a sound suited the contents of a video suggests that the rating scores did not differ between recall cued groups in either phase bin 2 or phase bin 4, nor between phase bins in each group. Neither was a significant interaction found between phase bin and cue condition.