Distinct Neural Representations of Content and Ordinal Structure in Auditory Sequence Memory

Ying Fan (范莹),^{1,2,3} Qiming Han (韩起明),^{2,3,4} Simeng Guo (郭思萌),⁵ and Huan Luo (罗欢)^{1,2,3}

¹School of Psychological and Cognitive Sciences, Peking University, Beijing, 100871, China, ²PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871, China, ³Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100871, China, ⁴Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, 100871, China, and ⁵Yuanpei College, Peking University, Beijing, 100871, China

Two forms of information, frequency (content) and ordinal position (structure), have to be stored when retaining a sequence of auditory tones in working memory (WM). However, the neural representations and coding characteristics of content and structure, particularly during WM maintenance, remain elusive. Here, in two EEG studies in human participants (both sexes), by transiently perturbing the “activity-silent” WM retention state and decoding the reactivated WM information, we demonstrate that content and structure are stored in a dissociative manner with distinct characteristics throughout WM process. First, each tone in the sequence is associated with two codes in parallel, characterizing its frequency and ordinal position, respectively. Second, during retention, a structural retrocue successfully reactivates structure but not content, whereas a following white noise triggers content but not structure. Third, structure representation remains stable, whereas content code undergoes a dynamic transformation through memory progress. Finally, the noise-triggered content reactivations during retention correlate with subsequent WM behavior. Overall, our results support distinct content and structure representations in auditory WM and provide an efficient approach to access the silently stored WM information in the human brain. The dissociation of content and structure could facilitate efficient memory formation via generalizing stable structure to new auditory contents.

Key words: auditory working memory; hidden state; multivariate decoding; reactivations; sequence structure

Significance Statement

In memory experiences, contents do not exist independently but are linked with each other via ordinal structure. For instance, recalling a piece of favorite music relies on correct ordering (sequence structure) of musical tones (content). How are the structure and content for an auditory temporally structured experience maintained in working memory? Here, by using impulse-response approach and time-resolved representational dissimilarity analysis on human EEG recordings in an auditory working memory task, we reveal that content and structure are stored in a dissociated way, which would facilitate efficient and rapid memory formation through generalizing stable structure knowledge to new auditory inputs.

Introduction

Memories are not stored in fragments; instead, multiple items or events are constantly linked and organized with each other according to certain relationships. For a sequence of items to be successfully retained in working memory (WM), two basic formats of information need to be encoded and maintained in the

brain: features that describe each item (content) and ordinal position of the item in the list (structure). For instance, making a phone call relies on correct assignment of ordinal labels to each digit. Retaining even a simplified version of auditory temporally structured experience (e.g., a tone sequence) necessitates storage of two types of code, content (e.g., frequency for each tone) and structure (e.g., ordinal position for each tone), in brain activities.

How does the human brain represent content and structure information in WM system? One possibility is that the two codes for a given item are represented in a combinational way; that is, each item is indexed by a specific content-structure combination (Komorowski et al., 2009; Naya et al., 2017; Kikumoto and Mayr, 2020). Alternatively, content and structure information could be represented independent of each other (Marshuetz, 2005; Bengio et al., 2013; Kalm and Norris, 2014; Higgins et al., 2017; Y. Liu et al., 2019), thereby facilitating efficient memory formation

Received Feb. 10, 2021; revised Apr. 25, 2021; accepted May 26, 2021.

Author contributions: Y.F. and H.L. designed the experiment; Y.F. and S.G. performed the experiment; Y.F. and Q.H. analyzed the data; Y.F. and H.L. wrote the paper.

This work was supported by National Natural Science Foundation of China 31930052 to H.L., and Beijing Municipal Science & Technology Commission Z181100001518002 to H.L.

The authors declare no competing financial interests.

Correspondence should be addressed to Huan Luo at huan.luo@pku.edu.cn.

<https://doi.org/10.1523/JNEUROSCI.0320-21.2021>

Copyright © 2021 the authors

(Behrens et al., 2018). Although amounts of previous studies have addressed the issue using different approaches in both animals and humans (e.g., Eichenbaum, 2014; Davachi and DuBrow, 2015; Kamiński and Rutishauser, 2020; Summerfield et al., 2020), most of the findings have been focused on the encoding period when the to-be-memorized stimuli are physically presented.

Meanwhile, when entering the WM retention period, the brain would reside in a relatively “activity-silent” state, especially for noninvasive electrophysiological recordings in humans. Specifically, WM information is difficult to be directly decoded from the neural recordings during the delay period and has been posited to be “silently” retained in synaptic weights of the network (e.g., Mongillo et al., 2008; Lewis-Peacock et al., 2012; Stokes, 2015; Rose et al., 2016; Masse et al., 2020). As a consequence, directly accessing neural representations of content and structure during retention and examining their WM behavioral relevance pose a huge challenge in the WM field, especially in human studies.

Here in two auditory sequence WM experiments, we examined how content and structure information are encoded and maintained in auditory WM, by using a time-resolved multivariate decoding approach on the EEG recordings. Crucially, as mentioned previously, since the recorded macroscopic activities tend to stay in an “activity-silent” state, we used an impulse-response approach (Wolff et al., 2017) to transiently perturb the neural network and then measure the subsequently reactivated WM information. Specifically, two triggering events were presented successively during retention: a structure retrocue and a white-noise auditory impulse. The former, as an abstract structure cue, is hypothesized to reactivate the stored structure representation (i.e., ordinal position), whereas the latter, a neutral white-noise auditory impulse, might trigger content information (i.e., tone frequency) by perturbing the auditory cortex where contents likely reside.

Our results demonstrate that content and ordinal structure of an auditory tone sequence are encoded and maintained in a dissociative manner with distinct characteristics. First, each presented tone during encoding is associated with two codes in parallel, characterizing its frequency and ordinal position, respectively. Second, during the “activity-silent” retention period, a structural retrocue successfully reactivates structure but not content, whereas a following white noise triggers content but not structure, implying their storage in different brain regions. Third, neural representation of structure information remains largely unchanged from encoding to retention, whereas content code undergoes dynamical transformation, signifying their distinct representational formats. Finally, the white-noise-triggered content reactivations during retention correlate with subsequent memory performance, confirming its genuine indexing of WM operations. Together, our results provide new evidence advocating a dissociated and distinct form of content-structure storage in auditory WM and constitute an efficient approach to directly access WM information in human brains.

Materials and Methods

Participants

Thirty (17 females and 13 males, mean age 23.4 years, range 19–27 years) and 20 (8 females and 12 males, mean age 19.7 years, range 18–25 years) healthy participants with normal or corrected-to-normal vision were recruited in Experiments 1 and 2, respectively, after providing written informed consent. Two participants in Experiment 2 were excluded because of low behavioral performance. Participants

received compensation for participation. The experiment was approved by the Departmental Ethical Committee of Peking University.

Apparatus and stimuli

The experiment stimuli were generated and controlled with MATLAB (the MathWorks) and Psychophysics Toolbox (Psychtoolbox-3) (Brainard, 1997). The visual stimuli were presented on a Display ++ LCD screen with a resolution of 1920 × 1080 pixels running at a refresh rate of 120 Hz. The distance between the screen and participants was fixed at 60 cm. Auditory stimuli were presented with a Sennheiser CX213 earphone through an RME Babyface pro external sound card. The intensity of all auditory stimuli was set at ~65 dB (62.1–67.3 dB) SPL. A USB keyboard was used for response collection.

Experimental procedure

Experiment 1. Each trial started with the presentation of a cross (0.9° visual angle), which stayed in the center of the screen throughout the entire trial, except during the retrocue’s presentation (see Fig. 1A). Participants were instructed to fixate at the central cross during the entire trial. After 700 ms, three memory tones with different frequencies that are pseudo-randomly selected from a uniform distribution of 6 frequencies (381, 538, 762, 1077, 1524, and 2155 Hz) were presented sequentially. The duration of each memory tone was 200 ms, with inter-stimulus interval of 700 ms. At 2000 ms after the offset of the third tone, a 200 ms visual retrocue (1.9° visual angle), a visual character (1, 2, or 3), was presented in the center of the screen, indicating which of the three memory tones would be tested later, with 100% validity. A 100 ms white-noise auditory impulse was then presented 2000 ms after the retrocue. Finally, after another 1500 ms interval, a 200 ms target auditory tone was presented. Participants made a judgment on whether the frequency of the target tone was higher or lower than the frequency of the cued memory tone, by pressing keys on the keyboard, without time limitation. The frequency of the target tone was either $2^{1/3}$ higher or lower, with equal probability, than the frequency of the cued memory tone. The next trial started with a 1500–2000 ms delay after participant made their responses. There were 1080 trials in total, which were separated into two sessions that were completed in 2 separate days, for each participant. Each session lasted ~3 h.

Experiment 2. Experiment 2 was the same as Experiment 1, except that participants were instructed to additionally report the ordinal position of the cued tone during recalling (see Fig. 5A). Furthermore, to control the temporal length of the experiment, there were other timing parameter adjustments; that is, the interval between the third auditory tone and the retrocue was set to 1500 ms, the interval between the retrocue and white noise to 1700 ms, and the interval between white noise and target tone to 1000 ms. Moreover, to prevent motor preparation during the maintaining period, the correspondence between particular ordinal positions and reaction keys is randomly set from six combinations in each trial. Same as in Experiment 1, there were 1080 trials in total, which were separated into two sessions being completed in 2 separate days, for each participant. Each session lasted ~3 h.

Experiment 3. Participants ($N = 19$, 12 females and 7 males; different participant groups from Experiments 1 and 2) did the same WM task as in Experiment 2 (see Fig. 5A), with or without presenting the white noise during the delay period. There were 180 trials for each condition (360 trials totally).

EEG acquisition and preprocessing

The EEG signals were acquired using a 64-channel actiCAP (Brain Products) and two BrainAmp amplifiers (Brain Products). The data were recorded through BrainVision Recorder software (Brain Products) at 500 Hz. Vertical electrooculogram was recorded by one additional electrode below the right eye. The impedances of all electrodes were kept to <10 k Ω . The EEG data were re-referenced to the average value across all channels, downsampled to 100 Hz, and bandpass filtered between 1 and 30 Hz. Independent component analysis was performed to remove eye-movement and other artifactual components. Data were epoched 500 ms before each trial’s onset to 700 ms after target tone’s offset.

A Experimental paradigm -- Exp 1 (N = 30)

B Memory tone (encoding) -- Exp 1 (N = 30)

C Topographies of channel-wise decoding (encoding) -- Exp 1 (N = 30)

Figure 1. Experimental paradigm and neural representations of content and structure during encoding period (Exp 1). **A**, Experimental paradigm (Exp 1). In each trial, three pure tones (first, second, and third Memory Tones) with different frequencies (selected from 381, 538, 762, 1077, 1524, and 2155 Hz) were serially presented, and participants were instructed to memorize both frequencies and ordinal positions of the three tones. During the maintaining period, a retrocue (1, 2, or 3) appeared to indicate which of the three tones would be tested at the end of the trial. Next, a completely task-irrelevant white-noise auditory impulse was presented. During the retrieval period, a target pure tone was presented, and participants needed to compare it with the cued memory tone (i.e., higher or lower in frequency). **B**, Neural representations of content (i.e., frequency) and structure (i.e., ordinal position) during encoding period. Top panels, Grand average (N = 30) neural representational dissimilarity (mean-centered) as a function of physical dissimilarity of the memory tone, for frequency (left) and ordinal position (right), as a function of time. Bottom panels, Grand average (N = 30, mean $\pm$ SEM) β values of the regression between neural representational dissimilarity and physical dissimilarity of memory tones (i.e., decoding performance), for frequency (left) and ordinal position (right), as a function of time. Gray bar on x axis represents memory tone presentation. Shaded area represents $\pm$ SEM across participants. *** $p < 0.001$. **C**, Topographies of channel-wise decoding results (frequency: 50–350 ms; ordinal position: 100–400 ms) for frequency (left) and ordinal position (right). Only correct trials were analyzed in each participant.

Epochs with extremely high noises by visual inspection were manually excluded from the following analyses.

Time-resolved multivariate decoding

Only the correct trials in each participant were used for further analyses, except the behavioral correlates analyses (see Figs. 4–6) during which the same number of correct and wrong trials in each participant were analyzed and compared (see details later in Materials and Methods).

Representational similarity analysis (RSA)

A time-resolved RSA was performed on the EEG data to evaluate the neural representation of frequency and ordinal position independently throughout the encoding and maintaining phases, at each time point for each participant. To remove the possible interference of slow trend, we first calculated the mean activity for each channel across all trials, and then the trial mean was smoothed by a 150 ms moving window and subtracted from the original data trial wisely (i.e., demean) (Grootevangers et

al., 2017). In addition, as the three pure tones were presented successively during encoding period, the difference of global field power (GFP) of these three tones might contribute to the ordinal position decoding analysis. To remove the possible confounding effects, we used a cross-validated confound regression method proposed by Snoek et al. (2019) to remove the variance that could be explained by the GFP levels from the encoding responses. Specifically, GFP levels for each ordinal position at each time point were first calculated and then entered into a linear regression model as predictor, with each channel's original activity at that time point as response variable. After estimating the contribution of GFP on EEG signals using this model for each channel and at each time point, variance associated with GFP was subtracted from the original data to obtain the corrected data. All the subsequent decoding analyses were conducted on the corrected data.

The RSA is based on both spatial and temporal information to achieve high signal-to-noise ratio (Grootswagers et al., 2017). Specifically, for RSA-based decoding at time point t , the values of all channels at the current time point t as well as those at previous time point $t - 1$ were all included as features ($64 \times 2 = 128$ in total) for a further eightfold cross-validation decoding approach. The Mahalanobis distance (Mahalanobis, 1936) in the neural spatiotemporal activities (i.e., 128 features) between each left-out test trial and the averaged, condition-specific activities over train trials was computed, with the covariance matrix estimated from all the train trials using a shrinkage estimator (Ledoit and Wolf, 2004). If the neural activities indeed contain information about certain feature (e.g., tone frequency, ordinal position), the more similar the two features are (less physical dissimilarity), the less Mahalanobis distance their associated neural activities would have (less neural representational dissimilarity). To further quantify their relationships, the physical differences were linearly regressed against the corresponding neural representational dissimilarity values, for each test trial. The mean β value of the regression across all test trials was then used to represent the decoding accuracy. This process was repeated 50 times with each containing a new random partition of data into 8 folds. The resulted decoding performance, smoothed with a Gaussian-weighted window (window length = 40 ms), was then averaged across the 50 partitions as the final decoding accuracy, at each time point and for each participant. The RSA was performed for frequency and ordinal position, independently, using the same EEG data but with different feature dimensions.

Spatial distribution

To illustrate the spatial distribution of RSA results for frequency and ordinal position, we performed a decoding analysis within certain time range (50–350 ms for tone; 100–400 ms for position code) on each EEG channel, using the time-resolved activities as features for multivariate decoding analysis (Kriegeskorte et al., 2006; van Ede et al., 2019). The time window is determined based on our decoding results (see Fig. 1B) as well as previous studies (e.g., Wolff et al., 2020).

Regressing out sensory factors

Notably, the ordinal position decoding results would contain the low-level sensory information carried by the retrocue stimuli (i.e., 1, 2, 3). To remove the confounding influences, we ran a control experiment during which participants were instructed to passively view the same retrocue stimuli (1, 2, 3), and their EEG responses were recorded. We then conducted a leave-one-participant-out decoding analysis on our main experiment data. Specifically, in each run, the response of one participant in the main experiment was used as test data, with the remaining participants' responses in the main experiment as train dataset 1 and those in the control experiment as train dataset 2. The neural response representational dissimilarity between the test data and train dataset 1 (containing both structural and sensory representations) and that between the test data and train dataset 2 (containing only sensory representations) were then calculated, respectively. Finally, the neural dissimilarity between the test data and train dataset 1 (containing both structural and sensory representations) was regressed according to two predictors, the ordinal positional dissimilarity (structure information) and the sensory neural dissimilarity (calculated between test data and

train dataset 2), with the former referring to the sensory-removed representational strength (see Fig. 2A, right, dotted line).

Behavioral relevance

For correct and wrong trial decoding analyses, the right trials were k -folded, where k was determined by the closest integer to the quotient of the total correct trial numbers and wrong trial numbers. For each partition, the $(k - 1)$ -fold correct trials were used as training data, and the left onefold correct trials and all wrong trials were used as testing data, respectively. This manipulation will ensure the same trial numbers for correct and wrong trials during testing. The above k -fold decoding procedure was further conducted 50 times, and averaged to obtain the final decoding time courses.

Cross-temporal generalization analysis

A cross-temporal generalization analysis was conducted to investigate whether the neural representations of frequency or ordinal position were similar for encoding and maintaining periods. We first selected the temporal range during the encoding period to serve as that for training data, based on their corresponding decoding performance during encoding (i.e., 90–130 ms and 140–180 ms after the onset of memory tone, for frequency and ordinal position, respectively). As shown in Figure 2, the neural representations of frequency and ordinal position only appeared after the white-noise stimulus and retrocue, respectively, during the delay period. Therefore, the cross-temporal generalization was performed only on the corresponding time ranges for these two features (i.e., 0–600 ms in reference to the onset of white-noise stimulus and retrocue) for frequency and ordinal position, respectively.

We used two types of multivariate decoding methods, RSA and lasso-regularized logistic regression model, for the cross-temporal generalization analysis. RSA was similar to previous analysis, except that the Mahalanobis distance was computed between responses during the encoding period (train data) and that during the maintaining period (test data). For lasso-regularized logistic regression model, we first trained classifier k ($k = \frac{1}{1 + e^{-x\beta + \alpha}}$), with x as 128 decoding features (64 channels over 2 time points) derived from EEG responses, vector β representing weights for each feature, and α as the intercept coefficient, to differentiate either ordinal positions or frequencies using the encoding-phase responses. The obtained classifiers were then applied on the maintaining-phase data. The output was next transformed to d' values, which represent the cross-temporal generalization decoding performance.

For both methods, the decoding results were further baseline-corrected (0–50 ms), repeated for 50 times, and then averaged. Finally, to remove possible influences from random effects, a control analysis was performed by shuffling test data labels and redoing the same cross-temporal generalization analysis for 50 times. The resulted shuffling results were then subtracted from the original results to get the final cross-temporal generalization results.

Statistical significance testing

Nonparametric sign-permutation test (customized analysis codes) (Maris and Oostenveld, 2007) was used for statistical test. Specifically, the sign of the decoding value of each participant at each time point was randomly flipped 100,000 times to obtain the null distribution, from which the p value was derived. Cluster-based permutation test was then conducted to correct multiple comparisons over time ($p < 0.05$). For clusters failing to pass the multiple comparison, the uncorrected results were reported.

Results

Experimental procedure and time-resolved multivariate decoding analysis

In Experiment 1, 30 human participants performed an auditory delayed-match-to-sample WM task while their 64-channel EEG activities were recorded. Each participant performed 1080 trials in total (6 h in 2 separate days). As shown in Figure 1A, in each

A Retrocue (maintaining) -- Exp 1 (N = 30) **B White noise (maintaining) -- Exp 1 (N = 30)**

Figure 2. Neural representations of content and structure during retention period (Exp 1). **A**, Neural representations of content (i.e., frequency) and structure (i.e., ordinal position) during retention period after retrocue. Top panels, Grand average ($N = 30$) neural representational dissimilarity (mean-centered) as a function of physical dissimilarity of the cued tone, for frequency (left) and ordinal position (right), as a function of time. Bottom panels, Grand average ($N = 30$, mean $\pm$ SEM) β values of the regression between neural representational dissimilarity and physical dissimilarity of cued tones (i.e., decoding performance), for frequency (left) and ordinal position (right), as a function of time. Gray bar on x axis represents retrocue presentation. Right, Dotted blue line indicates sensory-information-removed decoding strength for ordinal position, corrected. Notably, error bar was presented but relatively small. **B**, Same as in **A**, but during retention after white-noise auditory impulse. Gray bar on x axis represents white-noise presentation. Only correct trials were analyzed in each participant. Shaded area represents ± 1 SEM across participants. *** $p < 0.001$. ** $p < 0.05$. Solid line indicates corrected using cluster-based permutation test (cluster-forming threshold: $p < 0.05$).

trial, three pure tones with different frequencies that were pseudo-randomly selected from six fixed values (f_1 : 381Hz, f_2 : 538Hz, f_3 : 762Hz, f_4 : 1077Hz, f_5 : 1524Hz, f_6 : 2155Hz) were presented sequentially, and participants were required to memorize both the ordinal position and frequency of the three tones. During the delay period, a retrocue (1, 2, or 3) was first presented to indicate which of the three tones would be tested later, followed by a 100 ms white-noise auditory impulse. During the retrieval period, a target auditory tone was presented, and participants were instructed to determine whether its frequency was higher or lower than that of the cued tone. The frequency of the target tone was determined in a pretest so that the overall memory performance would be $\sim 75\%$. Indeed, the behavioral performance for the auditory WM task was 75.35% (SE = 1.00%).

A time-resolved RSA (Kriegeskorte et al., 2008) was performed on the EEG responses to evaluate the neural representation of frequency and ordinal position, respectively, throughout the encoding and maintaining phases, at each time point and in

each participant (see details in Materials and Methods). Notably, instead of binary decoding, the RSA here is based on a hypothesis that the neural representational similarity is proportional to the factor-of-interest similarity; that is, the neural response for f_1 (381Hz) and f_2 (538Hz) would be less dissimilar than that for f_1 (381Hz) and f_5 (1524Hz). To quantify the representational strength, we next calculated the linear regression (β) between the factor dissimilarity and the corresponding neural representational dissimilarity for frequency and ordinal position, respectively, that is, decoding one factor (e.g., frequency) while marginalizing the other factor (e.g., ordinal position).

Neural representations of structure and content during WM encoding

During the encoding period, each of the three tones could be characterized by two factors, that is, frequency ($f_1, f_2, f_3, f_4, f_5, f_6$) and ordinal position (first, second, third). We next

assessed the neural representations of content and structure, separately, from the same EEG response for each tone. Moreover, a cross-validated confound regression approach (Snoek et al., 2019) was used to regress out possible influences from GFP courses for the three tones on the multivariate decoding analyses.

Figure 1B (top panels) plots the time-resolved neural representational dissimilarity as a function of physical dissimilarity (y axis) after the onset of each tone, for frequency (left) and ordinal position (right). It is clear that, for both properties, the neural dissimilarity is minimum at center and gradually increases to the two sides when the physical difference becomes larger, supporting that the neural responses contain information for both content and structure. Figure 1B (bottom panel) shows the corresponding regression weights of the dissimilarity matrices, for frequency (β_{freq} , left) and ordinal position (β_{pos} , right), separately. Specifically, both the frequency representation β_{freq} (left; $p < 0.001$, corrected) and the ordinal position representation β_{pos} (right; $p < 0.001$, corrected) emerged shortly after the onset of each tone. Figure 1C plots the topographies of channel-wise decoding results for frequency (left) and ordinal position (right), respectively. Frequency decoding is associated with frontal-central and lateral channels, whereas structure representation is mainly located on parietal channels.

In summary, during the encoding period when a sequence of tones is presented to be retained in WM, each sound is signified by two separate neural codes that characterize its content (frequency) and structure (ordinal position), respectively, which likely originate from distinct brain areas.

Reactivations of structure and content information during “activity-silent” retention

After establishing the representations of content and structure information during the encoding phase, we next examined their maintenance during retention when the brain responses enter an “activity-silent” state. Crucially, the triggering events, retrocue and white noise, successfully reactivated WM information from the silent state, yet content (left) and structure (right) displayed distinct temporal profiles. Specifically, as shown in Figure 2A, the retrocue triggered the emergence of ordinal position information (right, blue line) ($p < 0.001$, corrected) but not that for tone frequency (left, red line; $p > 0.5$, corrected). In contrast, the subsequent white-noise auditory impulse (Fig. 2B) successfully activated the tone frequency representation (left, red line; 200–260 ms: $p = 0.017$, corrected) but not the ordinal position (right, blue line; $p > 0.359$, corrected). Thus, contents and ordinal structure are maintained in a dissociated way (i.e., being reactivated by different triggering events). The RSA decoding analysis was performed in reference to the memory-related information; for example, a trial with retrocue of 2 and $f4(1077\text{Hz})$ as the second memorized tone frequency would be labeled as 2 for ordinal position and $f4$ for frequency.

Importantly, since the retrocue stimuli are physically different (e.g., 1, 2, 3), the observed structure representation after the retrocue (Fig. 2A, right, blue solid line) would also contain the low-level sensory information. To address this issue, we further regressed the neural response dissimilarity according to two indexes: the ordinal positional dissimilarity (genuine structure information) and the dissimilarity between the neural response in the current experiment and that in a control experiment where participants viewed the identical retrocue stimuli (sensory information). This would allow us to remove the low-level sensory contributions from the structure decoding results. As shown

in Figure 2A, the sensory-removed decoding performance remained present (right, dotted blue line), supporting that the ordinal position representations are not just caused by sensory inputs. It might still be argued that there are other confounding factors carried by the retrocue stimuli, an issue we further addressed using an encoding-to-maintaining generalization analysis (Fig. 3).

Overall, during the “activity-silent” maintaining period, a structural retrocue successfully reactivates structure but not content information, whereas a following white-noise auditory impulse triggers WM content (to-be-recalled tone frequency) but not structure representation, implying their storage in different brain regions.

Distinct encoding-to-maintaining representational generalizations for structure and content

Having confirmed the neural representations during both encoding (Fig. 1B) and maintaining periods for structure (Fig. 2A, right; after retrocue) and content (Fig. 2B, left, after white noise), we next conducted an encoding-to-maintaining generalization analysis to estimate whether these two phases share similar coding formats. Put another way, if the neural codes for encoding (Fig. 1B) could be successfully generalized to those elicited during retention (Fig. 2A,B), this would support a stable WM coding, otherwise a dynamic coding transformation.

We first used the same RSA approach as before to examine the cross-temporal generalization (Fig. 3A), except that here the response dissimilarity was calculated between the encoding and retention activations. As shown in Figure 3B, only the structural information displayed a significant cross-temporal generalization (right; 230–280 ms: $p = 0.026$, corrected), but not for content (left; $p > 0.089$, corrected). To further verify the results, we also used a lasso-regularized logistic regression model to test the cross-temporal generalization. As shown in Figure 3C, likewise, only the structure information showed significant encoding-to-maintenance generalization (right; 90–570 ms: $p < 0.001$, corrected), but not for content (left; $p > 0.438$, corrected). Importantly, the successful cross-temporal generalization for ordinal position confirms the genuine structure information triggered by the retrocue during retention, since no visual cues (i.e., 1, 2, 3 stimuli) were presented during the encoding period yet the structural codes could still be generalized to that during retention (for other possible factors, see Discussion).

Together, structure and content are endowed with distinct representational transformation properties (i.e., structure representation remains stable from encoding to maintaining periods), while the content code undergoes a dynamic transformation, signifying their distinct representational formats (for full cross-time generalization results, see Fig. 6C).

Content reactivations during retention correlate with WM behavior

We next evaluated whether the neural representations of content and structure retained in WM, which could be reactivated by the retrocue and white noise, respectively, have any behavioral consequence. To this end, we performed the same RSA on correct trials and incorrect trials separately, in each participant. We ensured equal numbers of correct and incorrect trials to make a fair comparison (for details, see Materials and Methods). As shown in Figure 4A, the correct (dark red) and wrong trials (light red) displayed distinct frequency reactivations after the white noise (correct vs wrong; black asterisks; 210–250 ms: $p = 0.053$, corrected). Specifically, the correct-trial group (dark red) showed

A Encoding-to-maintaining representational generalization -- Exp 1 (N = 30) **B Cross-temporal generalization (representation similarity analysis) -- Exp 1 (N = 30)** **C Cross-temporal generalization (lasso-regularized logistic regression models) -- Exp 1 (N = 30)**

Figure 3. Encoding-to-maintaining representational generalization (Exp 1). **A**, Illustration of the cross-temporal generalization from encoding to maintaining periods for frequency (left) and ordinal position (right). Left, Frequency generalization; that is, encoding-phase response (within 90–130 ms) was used as train data to test the maintaining-phase responses at each time point after white-noise auditory impulse. Right, Ordinal position generalization; that is, encoding-phase response (within 140–180 ms) was used as train data to test the maintaining-phase responses at each time point after retrocue. **B**, Grand average ($N = 30$, mean $\pm$ SEM) cross-temporal decoding generalization results using RSA as a function of time during maintaining, for frequency (left) and ordinal position (right). Gray bar on x axis represents white-noise (left) and retrocue (right). **C**, Same as in **B**, but using lasso-regularized logistic regression models. Only correct trials were analyzed in each participant. Shaded area represents ± 1 SEM across participants. *** $p < 0.001$. ** $p < 0.05$. Solid line indicates corrected using cluster-based permutation test (cluster-forming threshold: $p < 0.05$).

significant frequency decoding performance (200–260 ms: $p = 0.018$, corrected), whereas the wrong-trial group (light red) did not. Interestingly, the wrong-trial group showed a trend of late reactivation (440–460 ms: $p = 0.178$, corrected; maximum $p = 0.049$, uncorrected).

Meanwhile, the structure reactivations (i.e., encoding-to-maintaining generalization, given its genuine indexing of structure information independent of retrocue stimulus) displayed no behavioral relevance (i.e., correct and wrong trials showed comparable structure reactivations) (RSA, Fig. 4B, $p > 0.138$, corrected; Lasso-regularized logistic regression model, Fig. 4C, $p > 0.106$, corrected). Furthermore, the neural representation during the encoding period did not show behavioral correlates

for both frequency (Fig. 4D, $p > 0.287$, corrected) and ordinal position (Fig. 4E, $p > 0.5$, corrected), suggesting that the deteriorating WM performance in wrong trials was not attributable to encoding failure.

Together, content information (i.e., tone frequency) maintained in WM, which could be assessed using a neutral white-noise auditory impulse during retention, covaries with subsequent WM recalling performance, further supporting its genuine indexing of WM operations.

Experiment 2: task control

In Experiment 1, participants were instructed to retain a list of auditory tones and to recall the tone frequency according to the

Exp 1 (N = 30)

A Frequency representation (maintaining)

D Frequency representation (encoding)

B Ordinal position generalization (RSA)

E Ordinal position representation (encoding)

C Ordinal position generalization (Regression Models)

Figure 4. Behavioral correlates of content and structure representations in auditory WM (Exp 1). **A**, Grand average ($N = 30$, mean $\pm$ SEM) frequency decoding performance during maintaining period after white-noise auditory impulse, for correct (dark red) and wrong (light red) trials. Gray bar on x axis represents white-noise presentation. **B**, Grand average ($N = 30$, mean $\pm$ SEM) structure decoding performance during maintaining period after retrocue (i.e., encoding-to-maintaining generalization), for correct (dark blue) and wrong (light blue) trials. Gray bar on x axis represents retrocue presentation. **C**, Same as in **B**, but using lasso-regularized logistic regression models. **D**, Same as in **A**, but during encoding period. Gray bar on x axis represents memory tone presentation. **E**, Same as in **B**, but during encoding period. Gray bar on x axis represents memory tone presentation. The decoding analyses were performed on correct and wrong trials with equal trial numbers. Shaded area represents ± 1 SEM across participants. *** $p < 0.001$. ** $p < 0.05$. Solid line indicates corrected using cluster-based permutation test (cluster-forming threshold: $p < 0.05$). Dotted line indicates uncorrected.

retrocue (first, second, third). Meanwhile, since the recalling task only tested the content, participants might have discarded the structure information after the retrocue. This would constitute an alternative interpretation for why the white-noise impulse failed to trigger structure representations (Fig. 2*B*, right). To address this possibility, we performed Experiment 2 ($N = 18$),

during which the participants did exactly the same task as Experiment 1 but needed to additionally recall the position of the memorized tone (Fig. 5*A*).

The behavioral accuracies for frequency and ordinal position were 75.17% (SE = 1.08%) and 98.80% (SE = 0.22%), respectively. As shown in Figure 5, Experiment 2 largely replicated the

A Experimental paradigm -- Exp 2 (N = 18) **B Retrocue (maintaining) -- Exp 2 (N = 18)** **C White noise (maintaining) -- Exp 2 (N = 18)** **D Cross-temporal generalization (representation similarity analysis) -- Exp 2 (N = 18)**

Figure 5. Control EEG experiment (Exp 2). **A**, Experiment 2 used the same design as Experiment 1 (see Fig. 1A), except explicitly requiring participant to hold the ordinal position in WM. Specifically, during the recalling period, a target pure tone was presented, and participants needed to compare it with the cued memory tone, that is, higher or lower in frequency (same as Experiment 1), and reported the ordinal position of the cued memory tone (first, second, third) by pressing corresponding keys (A, S, and D). **B**, Neural representations of content (i.e., frequency) and structure (i.e., ordinal position) during maintaining period after retrocue. Grand average ($N = 18$, mean $\pm$ SEM) β values of the regression between neural representational dissimilarity and physical dissimilarity of cued tones, for frequency (left) and ordinal position (right), as a function of time. **C**, Left, Grand average ($N = 18$, mean $\pm$ SEM) frequency decoding

results of Experiment 1, thus excluding the task interpretation. Specifically, the retrocue elicited structure (Fig. 5B, right; 60–600 ms: $p = 0.003$, corrected) but not frequency (Fig. 5B, left; $p > 0.337$, corrected), whereas the white noise reactivated content (Fig. 5C, left, dark red; 250–320 ms: $p = 0.006$, corrected) but still failed to drive structure (Fig. 5C, right, $p > 0.350$, corrected). Moreover, same as in Experiment 1, the cross-temporal generalization results showed a significant encoding-to-maintaining generalization for structure (Fig. 5D, right, dark blue; 140–180 ms: $p = 0.046$, corrected) but not for content (Fig. 5D, left, $p > 0.251$, corrected). Finally, also consistent with Experiment 1, correct trials showed better content reactivations than wrong trials (Fig. 5C, left, black star, 250–280 ms: $p = 0.054$, corrected), but not for ordinal position (Fig. 5D, right, $p > 0.158$, corrected). Together, the new control experiment fully replicated previous results, supporting that the lack of structure reactivation after the white-noise impulse is not because of task in Experiment 1.

We combined Experiments 1 and 2 ($N = 48$), given that they shared the same design through encoding and retention and showed similar response patterns. Consistently, the white noise triggered stronger content reactivation for correct than incorrect trials (Fig. 6A, black, 210–270 ms: $p = 0.011$, corrected), whereas the retrocue did not (Fig. 6B, $p > 0.148$, corrected). Intriguingly, there was still a trend of late content reactivation for incorrect trials (430–450 ms: $p = 0.179$, corrected; maximum $p = 0.045$, uncorrected), which was even marginally higher than that for correct trials (430–450 ms: $p = 0.178$, corrected; maximum $p = 0.045$, uncorrected), implying that the WM content in wrong trials might not be lost completely but tends to be maintained in a less excitable latent state of WM network (see Discussion).

Figure 6C plots the full cross-time encoding-to-maintaining representational generalization results, for content and structure, respectively. It is clear that no significant frequency generalization was detected at all (Fig. 6C, left), ensuring that the lack of coding generalization for content (Fig. 6A) is not because of the selection of specific temporal windows. In contrast, structure information revealed significant coding generalization at relatively stable times after retrocue (right panel), confirming the stable coding for structure.

Experiment 3: control behavioral experiment

Finally, we conducted a control behavioral experiment ($N = 19$; different participants from Experiments 1 and 2) to test whether the added white-noise PING stimulus during retention would influence the WM performance. Participants did the same WM task as in Experiment 2 (Fig. 5A), with or without presenting the white noise during the delay period. As plotted in Figure 6D, the two conditions (with noise, without noise) showed no WM behavioral difference in either accuracy or reaction time on content and structure, supporting that the presented white-noise PING during retention would not interfere with WM performance. Therefore, consistent with previous findings (Wolff et al., 2017,

2020), the PING stimulus serves as a neutral probe to access information retained in WM system.

Discussion

As we learn about the world, memories are formed not only for individual events but also for their relational structure. Here, we demonstrate that content (tone frequency) and structure (ordinal position) of a tone sequence are encoded and stored in a distinct way in human auditory WM. Each item is signified by two separate neural codes, content and structure, which could be subsequently reactivated from the “activity-silent” state during WM retention, by white-noise auditory impulse and retrocue, respectively. Furthermore, content and structure demonstrate distinct temporal properties throughout memory process (i.e., stable representation for structure and dynamical coding for content). Importantly, the content reactivations during retention are further correlated with WM recalling performance. Overall, the human brain extracts and separates structure and content information from auditory inputs and maintains them in distinct formats in the WM network. Our work also provides an efficient approach to access the “silently” stored WM information in the human brain.

It has long been posited that structure information that characterizes the abstract relationship between objects serves as a primary rule to organize fragmented contents and influences perception: that is, Gestalt principle (Wagemans et al., 2012), memory (DuBrow and Davachi, 2013; Gershman et al., 2013; Davachi and DuBrow, 2015), and learning (Luyckx et al., 2019). How is the structure information implemented in the brain? One typical example is the hippocampus neurons that are sensitive to specific spatial location or time bin that are independent of the attached contents, signifying their coding of spatial (O’Keefe and Dostrovsky, 1971; O’Keefe and Nadel, 1978; Muller, 1996; Eichenbaum et al., 1999), temporal (Manns et al., 2007; Paz et al., 2010; MacDonald et al., 2011; Eichenbaum, 2014; Heusser et al., 2016; Tsao et al., 2018), and abstract structures even in nonspatial tasks, i.e., cognitive map (Aronov et al., 2017; Garvert et al., 2017; Schuck and Niv, 2019). Parietal regions are also involved in structural representation and learning (Feigenson et al., 2004; Summerfield et al., 2020). Here we focused on the ordinal information of a tone sequence, a typical and important relational structure to sort contents in auditory experience. Indeed, neural coding of ordinal position has been found in both animal recordings (Averbeck et al., 2006; Naya and Suzuki, 2011; Kraus et al., 2013; Crowe et al., 2014; Taxidis et al., 2020) and neuroimaging studies (Roberts et al., 2013; Hsieh et al., 2014; DuBrow and Davachi, 2016; Baldassano et al., 2017; Rajji et al., 2017; Carpenter et al., 2018; Huang et al., 2018; Kikumoto and Mayr, 2018; X. L. Liu et al., 2020). Our results are thus consistent with previous findings and further expand to auditory sequence memory, particularly during the maintaining period.

The successful encoding-to-maintaining representational generalization for ordinal position serves as crucial evidence for structure representation in WM by ruling out some alternative interpretations. For example, the structure code for each tone during encoding might arise from repetition suppression. Meanwhile, the encoding-to-maintaining representational generalization excludes this possibility since the retrocue is presented only once and the structure reactivation in turn could not be because of repetition suppression during retention. Furthermore, subjects might use a verbal label strategy for each ordinal position, which constitutes another explanation for the observed

← performance after white noise, for correct (dark red) and wrong (light red) trials. Right, Grand average ordinal position decoding performance after white noise. D, Grand average ($N = 18$, mean $\pm$ SEM) encoding-to-maintaining generalization results for frequency (left) and ordinal position (right) for correct (dark blue) and wrong (light blue) trials, with equal numbers of correct and wrong trials in each participant. Shaded area represents ± 1 SEM across participants. *** $p < 0.001$. ** $p < 0.05$. Solid line indicates corrected using cluster-based permutation test (cluster-forming threshold: $p < 0.05$).

A White noise (maintaining) -- Exp 1 & 2 (N = 48) **B Ordinal position generalization -- Exp1 & 2** **C Cross-temporal generalization (representation similarity analysis) -- Exp 1 & 2 (N = 48)** **D White noise does not interfere with memory performance -- Control Exp (N = 19)**

Figure 6. Combination of Experiments 1 and 2 (Exp 1 and Exp 2), cross-time encoding-to-maintaining generalization, and behavioral control experiment. **A**, Grand average ($N=48$, mean $\pm$ SEM) frequency decoding performance after white noise, for correct (dark red) and wrong (light red) trials, by combining Experiments 1 and 2. **B**, Grand average ($N=48$, mean $\pm$ SEM) encoding-to-maintaining generalization results for ordinal position after retrocue, for correct (dark blue) and wrong (light blue) trials, by combining Experiments 1 and 2, with equal numbers of correct and wrong trials in each participant. Shaded area represents ± 1 SEM across participants. $**p < 0.05$. Solid line indicates corrected using cluster-based permutation test (cluster-forming threshold: $p < 0.05$). Dotted line indicates uncorrected. **C**, Grand average ($N=48$) cross-time encoding-to-maintaining generalization for frequency (left) and ordinal position (right), by combining Experiments 1 and 2. Areas enclosed by solid red lines represent significant decoding (cluster correction, cluster-forming threshold: $p < 0.05$; $p < 0.05$). Dotted red lines indicate the time window selected in previous generalization analyses (see Figs. 3B,C, 4B,C, 5D, 6B). **D**, Control behavioral experiment testing whether the presentation of white-noise PING stimulus during retention would affect memory performance (with noise vs without noise). Each dot represents individual result. +, Outliers.

structural code. However, since the task is to memorize the frequencies of three tones, subjects reported that they were actually perceiving and rehearsing a sequence of three tones instead of three verbal labels. In addition, animal studies also demonstrate neural representations for temporal order (Fortin et al., 2002; Manns et al., 2007; Berdyeva and Olson, 2010; Rangel et al., 2014; Allen et al., 2016), which obviously involve no verbal label processing. Together, we posit that the structure code observed here reflects a genuine, abstract representation of ordinal position in sequence WM that are shared across species.

Different from visual studies (Foster et al., 2016; Fukuda et al., 2016), only a few studies have assessed the neural correlates of auditory WM (Luo et al., 2013; Kumar et al., 2016; Albouy et al., 2017; Weisz et al., 2020; Wolff et al., 2020). Here, by combining RSA-based decoding and impulse-response approach, we demonstrate that a neutral white-noise auditory impulse could successfully reactivate the tone frequency information “silently” held in auditory WM. This is consistent with the hidden-state WM view, which posits that information is stored in synaptic weights via short-term neural plasticity principles (Mongillo et al., 2008; Stokes, 2015). As a result, a transient disturbance of the neural network (e.g., a flash or a white noise) would endow the neural assemblies that maintain WM information in synaptic weights with larger probability to be activated (Stokes, 2015). In other words, the white-noise here serves as a “neutral” probe to assess the information that has already been retained in WM and would not modify WM contents. Nevertheless, not all the triggering events could efficiently reactivate contents (e.g., the retrocue failed to trigger content information), implying that content and structure might be stored in different regions and have different characteristics. Auditory content, presumably retained in sensory cortex, is sensitive to auditory perturbation, whereas structure information might be maintained in high-level areas (e.g., parietal region) (Buettner and Walsh, 2009; Parkinson et al., 2014; Summerfield et al., 2020) and frontal cortex (Ninokura et al., 2003; Berdyeva and Olson, 2010; Hsieh et al., 2011; Naya et al., 2017). Notably, our results advocating the “activity-silent” view are based on multivariate decoding analyses in noninvasive EEG recordings. Studies using new techniques or animal recordings will deepen our understanding of the interesting view.

The white-noise-elicited content reactivation during retention correlates with subsequent recalling performance, further confirming its genuine indexing of content maintenance in auditory WM. Interestingly, a trend of late content reactivation for incorrect trials suggests that content information might not be completely lost but still retained in WM, yet in a less excitable latent state (i.e., being triggered in a late and less robust way). A recent visual WM study reveals that top-down attention could modulate the latent states of WM items; that is, the most task-relevant one would be in a more excitable state and tends to be reactivated earlier by TMS perturbation (Rose et al., 2016). Moreover, computational modeling that incorporates competition between items also predicts late reactivation for weakly stored WM items (Mongillo et al., 2008). Therefore, in addition to reactivation strength, the latency of content reactivations might also serve as a potential index to assess subsequent WM performance.

What would be the benefit of the dissociated content and structure representation in WM system? An apparent advantage is that this would allow rapid and versatile transfer of the structure information to new contents (Tse et al., 2007; Friston and Buzsáki, 2016; Kaplan et al., 2017; Behrens et al., 2018; Whittington et al., 2018). Visual WM studies also reveal that

different attributes (e.g., color, shape, or orientation) of a single object are coded independently (Bays et al., 2011; Fougner and Alvarez, 2011; Cowan et al., 2013). Back to the current experiment, all trials share the fixed sequence structure (e.g., 3-tone sequence), and it would therefore be more efficient for the brain to separately encode and maintain the ordinal position information and allocate tone information anew in each trial to the corresponding ordinal positions.

Finally, structure and content also show different coding dynamics (i.e., stable and dynamic representations, respectively). The transformed code for WM contents has been widely found (e.g., Meyers et al., 2008; Barak et al., 2010; Lundqvist et al., 2016, 2018; Sprague et al., 2016; Parthasarathy et al., 2017, 2019; Spaak et al., 2017; Trübetschek et al., 2017; Wolff et al., 2017, 2020; Quentin et al., 2019; Rademaker et al., 2019; Yue et al., 2019; Kamiński and Rutishauser, 2020; Yu et al., 2020). Notably, these studies do not always use the PING approach but still observe the transformed representation for WM contents, thereby ruling out possibilities that the dynamic code for content is because of the presentation of PING stimulus. Moreover, the control behavioral experiment (Fig. 6D) suggests that the PING stimulus serves as a neutral probe to access WM representation and would not interfere with WM performance. Possible interpretations for the dynamic coding are that WM content is coded in a way that is optimized for subsequent behavioral demands (Myers et al., 2017; Panichello and Buschman, 2020) or transformed into a subspace to resist distractions from new inputs (Murray et al., 2017; Libby and Buschman, 2019). In contrast, the ordinal structure displays a stable representation over memory course, also in line with previous findings (Walsh, 2003; Kalm and Norris, 2017; Luyckx et al., 2019; Bernardi et al., 2020). Stable structure coding would be advantageous for memory generalization and formation (i.e., rapid implementation of stable structure to dynamic contents).

Together, content (tone frequency) and structure (ordinal position) are the two basic formats of information to be maintained in WM. Our findings demonstrate that they are encoded and stored in a largely dissociated way and display distinct representational characteristics, which echo their varied functions in memory formation: detailed and dynamic contents versus abstract and stable structure.

References

- Albouy P, Weiss A, Baillet S, Zatorre RJ (2017) Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron* 94:193–206.e195.
- Allen TA, Salz DM, McKenzie S, Fortin NJ (2016) Nonspatial sequence coding in CA1 neurons. *J Neurosci* 36:1547–1563.
- Aronov D, Nevers R, Tank DW (2017) Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature* 543:719–722.
- Averbeck BB, Sohn JW, Lee D (2006) Activity in prefrontal cortex during dynamic selection of action sequences. *Nat Neurosci* 9:276–282.
- Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA (2017) Discovering event structure in continuous narrative perception and memory. *Neuron* 95:709–721.e705.
- Barak O, Tsodyks M, Romo R (2010) Neuronal population coding of parametric working memory. *J Neurosci* 30:9424–9430.
- Bays PM, Wu EY, Husain M (2011) Storage and binding of object features in visual working memory. *Neuropsychologia* 49:1622–1631.
- Behrens TE, Muller TH, Whittington JC, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018) What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100:490–509.
- Bengio Y, Mesnil G, Dauphin Y, Rifai S (2013) Better mixing via deep representations. In *International conference on machine learning*, pp 552–560. PMLR, Atlanta, Georgia, USA.

- Berdyyeva TK, Olson CR (2010) Rank signals in four areas of macaque frontal cortex during selection of actions and objects in serial order. *J Neurophysiol* 104:141–159.
- Bernardi S, Benna MK, Rigotti M, Munuera J, Fusi S, Salzman CD (2020) The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* 183:954–967.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Buetti D, Walsh V (2009) The parietal cortex and the representation of time, space, number and other magnitudes. *Philos Trans R Soc Lond B Biol Sci* 364:1831–1840.
- Carpenter AF, Baud-Bovy G, Georgopoulos AP, Pellizzer G (2018) Encoding of serial order in working memory: neuronal activity in motor, premotor, and prefrontal cortex during a memory scanning task. *J Neurosci* 38:4912–4933.
- Cowan N, Blume CL, Sauls JS (2013) Attention to attributes and objects in working memory. *J Exp Psychol Learn Mem Cogn* 39:731–747.
- Crowe DA, Zarco W, Bartolo R, Merchant H (2014) Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *J Neurosci* 34:11972–11983.
- Davachi L, DuBrow S (2015) How the hippocampus preserves order: the role of prediction and context. *Trends Cogn Sci* 19:92–99.
- DuBrow S, Davachi L (2013) The influence of context boundaries on memory for the sequential order of events. *J Exp Psychol Gen* 142:1277–1286.
- DuBrow S, Davachi L (2016) Temporal binding within and across events. *Neurobiol Learn Mem* 134:107–114.
- Eichenbaum H (2014) Time cells in the hippocampus: a new dimension for mapping memories. *Nat Rev Neurosci* 15:732–744.
- Eichenbaum H, Dudchenko P, Wood E, Shapiro M, Tanila H (1999) The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron* 23:209–226.
- Feigenson L, Dehaene S, Spelke E (2004) Core systems of number. *Trends Cogn Sci* 8:307–314.
- Fortin NJ, Agster KL, Eichenbaum HB (2002) Critical role of the hippocampus in memory for sequences of events. *Nat Neurosci* 5:458–462.
- Foster JJ, Sutterer DW, Serences JT, Vogel EK, Awh E (2016) The topography of alpha-band activity tracks the content of spatial working memory. *J Neurophysiol* 115:168–177.
- Fougnie D, Alvarez GA (2011) Object features fail independently in visual working memory: evidence for a probabilistic feature-store model. *J Vis* 11:10.1167/11.12.3.3.
- Friston K, Buzsáki G (2016) The functional anatomy of time: what and when in the brain. *Trends Cogn Sci* 20:500–511.
- Fukuda K, Kang MS, Woodman GF (2016) Distinct neural mechanisms for spatially lateralized and spatially global visual working memory representations. *J Neurophysiol* 116:1715–1727.
- Garvert MM, Dolan RJ, Behrens TE (2017) A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *Elife* 6:e17086.
- Gershman SJ, Schapiro AC, Hupbach A, Norman KA (2013) Neural context reinstatement predicts memory misattribution. *J Neurosci* 33:8590–8595.
- Grootswagers T, Wardle SG, Carlson TA (2017) Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J Cogn Neurosci* 29:677–697.
- Heusser AC, Poeppel D, Ezzyat Y, Davachi L (2016) Episodic sequence memory is supported by a theta–gamma phase code. *Nat Neurosci* 19:1374–1380.
- Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S, Lerchner A (2017) β -VAE: Learning basic visual concepts with a constrained variational framework. Paper presented at the International Conference on Learning Representations, Toulon, France.
- Hsieh LT, Ekstrom AD, Ranganath C (2011) Neural oscillations associated with item and temporal order maintenance in working memory. *J Neurosci* 31:10803–10810.
- Hsieh LT, Gruber MJ, Jenkins LJ, Ranganath C (2014) Hippocampal activity patterns carry information about objects in temporal context. *Neuron* 81:1165–1178.
- Huang Q, Jia J, Han Q, Luo H (2018) Fast-backward replay of sequentially memorized items in humans. *Elife* 7:e35164.
- Kalm K, Norris D (2014) The representation of order information in auditory-verbal short-term memory. *J Neurosci* 34:6879–6886.
- Kalm K, Norris D (2017) A shared representation of order between encoding and recognition in visual short-term memory. *Neuroimage* 155:138–146.
- Kamiński J, Rutishauser U (2020) Between persistently active and activity-silent frameworks: novel vistas on the cellular basis of working memory. *Ann NY Acad Sci* 1464:64–75.
- Kaplan R, Schuck NW, Doeller CF (2017) The role of mental maps in decision-making. *Trends Neurosci* 40:256–259.
- Kikumoto A, Mayr U (2018) Decoding hierarchical control of sequential behavior in oscillatory EEG activity. *eLife* 7:e38550.
- Kikumoto A, Mayr U (2020) Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. *Proc Natl Acad Sci USA* 117:10603–10608.
- Komorowski RW, Manns JR, Eichenbaum H (2009) Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. *J Neurosci* 29:9918–9929.
- Kraus BJ, Robinson RJ II, White JA, Eichenbaum H, Hasselmo ME (2013) Hippocampal “time cells”: time versus path integration. *Neuron* 78:1090–1101.
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103:3863–3868.
- Kriegeskorte N, Mur M, Bandettini PA (2008) Representational similarity analysis-connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD (2016) A brain system for auditory working memory. *J Neurosci* 36:4492–4505.
- Ledoit O, Wolf M (2004) Honey, I shrunk the sample covariance matrix. *JPM* 30:110–119.
- Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR (2012) Neural evidence for a distinction between short-term memory and the focus of attention. *J Cogn Neurosci* 24:61–79.
- Libby A, Buschman TJ (2019) Rotational dynamics reduce interference between sensory and memory representations. *bioRxiv* 641159.
- Liu Y, Dolan RJ, Kurth-Nelson Z, Behrens TE (2019) Human replay spontaneously reorganizes experience. *Cell* 178:640–652.e614.
- Liu XL, Ranganath C, Hsieh LT, Hurtado M, Niendam TA, Lesh TA, Carter CS, Ragland JD (2020) Task-specific disruptions in theta oscillations during working memory for temporal order in people with schizophrenia. *J Cogn Neurosci* 32:2117–2130.
- Lundqvist M, Rose J, Herman P, Brincat SL, Buschman TJ, Miller EK (2016) Gamma and beta bursts underlie working memory. *Neuron* 90:152–164.
- Lundqvist M, Herman P, Miller EK (2018) Working memory: delay activity, yes! Persistent activity? Maybe not. *J Neurosci* 38:7013–7019.
- Luo H, Tian X, Song K, Zhou K, Poeppel D (2013) Neural response phase tracks how listeners learn new acoustic representations. *Curr Biol* 23:968–974.
- Luyckx F, Nili H, Spitzer B, Summerfield C (2019) Neural structure mapping in human probabilistic reward learning. *Elife* 8:e42816.
- MacDonald CJ, Lepage KQ, Eden UT, Eichenbaum H (2011) Hippocampal “time cells” bridge the gap in memory for discontinuous events. *Neuron* 71:737–749.
- Mahalanobis PC (1936) On the generalised distance in statistics. *Proceedings of the National Institute of Science (Calcutta)*, vol 2, pp 49–55.
- Manns JR, Howard MW, Eichenbaum H (2007) Gradual changes in hippocampal activity support remembering the order of events. *Neuron* 56:530–540.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- Marshuetz C (2005) Order information in working memory: an integrative review of evidence from brain and behavior. *Psychol Bull* 131:323–339.
- Masse NY, Rosen MC, Freedman DJ (2020) Reevaluating the role of persistent neural activity in short-term memory. *Trends Cogn Sci* 24:242–258.
- Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T (2008) Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J Neurophysiol* 100:1407–1419.
- Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. *Science* 319:1543–1546.
- Muller R (1996) A quarter of a century of place cells. *Neuron* 17:813–822.
- Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, Wang XJ (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci USA* 114:394–399.

- Myers NE, Stokes MG, Nobre AC (2017) Prioritizing information during working memory: beyond sustained internal attention. *Trends Cogn Sci* 21:449–461.
- Naya Y, Chen H, Yang C, Suzuki WA (2017) Contributions of primate prefrontal cortex and medial temporal lobe to temporal-order memory. *Proc Natl Acad Sci USA* 114:13555–13560.
- Naya Y, Suzuki WA (2011) Integrating what and when across the primate medial temporal lobe. *Science* 333:773–776.
- Ninokura Y, Mushiake H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J Neurophysiol* 89:2868–2873.
- O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34:171–175.
- O'Keefe J, Nadel L (1978) *The hippocampus as a cognitive map*. Oxford: Clarendon.
- Panichello MF, Buschman TJ (2020) Selective control of working memory in prefrontal, parietal, and visual cortex. *bioRxiv* 2020.04.07.030718.
- Parkinson C, Liu S, Wheatley T (2014) A common cortical metric for spatial, temporal, and social distance. *J Neurosci* 34:1979–1987.
- Parthasarathy A, Herikstad R, Bong JH, Medina FS, Libedinsky C, Yen SC (2017) Mixed selectivity morphs population codes in prefrontal cortex. *Nat Neurosci* 20:1770–1779.
- Parthasarathy A, Tang C, Herikstad R, Cheong LF, Yen SC, Libedinsky C (2019) Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nat Commun* 10:1–11.
- Paz R, Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I (2010) A neural substrate in the human hippocampus for linking successive events. *Proc Natl Acad Sci USA* 107:6046–6051.
- Quentin R, King JR, Sallard E, Fishman N, Thompson R, Buch ER, Cohen LG (2019) Differential brain mechanisms of selection and maintenance of information during working memory. *J Neurosci* 39:3728–3740.
- Rademaker RL, Chunharas C, Serences JT (2019) Coexisting representations of sensory and mnemonic information in human visual cortex. *Nat Neurosci* 22:1336–1344.
- Rajji TK, Zomorodi R, Barr MS, Blumberger DM, Mulsant BH, Daskalakis ZJ (2017) Ordering information in working memory and modulation of gamma by theta oscillations in humans. *Cereb Cortex* 27:1482–1490.
- Rangel LM, Alexander AS, Aimone JB, Wiles J, Gage FH, Chiba AA, Quinn LK (2014) Temporally selective contextual encoding in the dentate gyrus of the hippocampus. *Nat Commun* 5:1–9.
- Roberts BM, Hsieh LT, Ranganath C (2013) Oscillatory activity during maintenance of spatial and temporal information in working memory. *Neuropsychologia* 51:349–357.
- Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyering EE, Postle BR (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354:1136–1139.
- Schuck NW, Niv Y (2019) Sequential replay of nonspatial task states in the human hippocampus. *Science* 364:eaaw5181.
- Snoek L, Miletić S, Scholte HS (2019) How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage* 184:741–760.
- Spaak E, Watanabe K, Funahashi S, Stokes MG (2017) Stable and dynamic coding for working memory in primate prefrontal cortex. *J Neurosci* 37:6503–6516.
- Sprague TC, Ester EF, Serences JT (2016) Restoring latent visual working memory representations in human cortex. *Neuron* 91:694–707.
- Stokes MG (2015) 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci* 19:394–405.
- Summerfield C, Luyckx F, Sheahan H (2020) Structure learning and the posterior parietal cortex. *Prog Neurobiol* 184:101717.
- Taxidis J, Pnevmatikakis EA, Dorian CC, Mylavarapu AL, Arora JS, Samadian KD, Hoffberg EA, Golshani P (2020) Differential emergence and stability of sensory and temporal representations in context-specific hippocampal sequences. *Neuron* 108:984–998.
- Trübetschek D, Marti S, Ojeda A, King JR, Mi Y, Tsodyks M, Dehaene S (2017) A theory of working memory without consciousness or sustained activity. *Elife* 6:e23871.
- Tsao A, Sugar J, Lu L, Wang C, Knierim JJ, Moser MB, Moser EI (2018) Integrating time from experience in the lateral entorhinal cortex. *Nature* 561:57–62.
- Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RG (2007) Schemas and memory consolidation. *Science* 316:76–82.
- Van Ede F, Chekroud SR, Stokes MG, Nobre AC (2019) Concurrent visual and motor selection during visual working memory guided action. *Nat Neurosci* 22:477–483.
- Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M, von der Heydt R (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol Bull* 138:1172–1217.
- Walsh V (2003) A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn Sci* 7:483–488.
- Weisz N, Kraft NG, Demarchi G (2020) Auditory cortical alpha/beta desynchronization prioritizes the representation of memory items during a retention period. *Elife* 9:e55508.
- Whittington JC, Muller TH, Mark S, Barry C, Behrens TE (2018) Generalisation of structural knowledge in the hippocampal-entorhinal system. *arXiv:1805.09042*.
- Wolff MJ, Jochim J, Akyürek EG, Stokes MG (2017) Dynamic hidden states underlying working-memory-guided behavior. *Nat Neurosci* 20:864–871.
- Wolff MJ, Kandemir G, Stokes MG, Akyürek EG (2020) Unimodal and bimodal access to sensory working memories by auditory and visual impulses. *J Neurosci* 40:671–681.
- Yu Q, Teng C, Postle BR (2020) Different states of priority recruit different neural representations in visual working memory. *PLoS Biol* 18:e3000769.
- Yue Q, Martin RC, Hamilton AC, Rose NS (2019) Non-perceptual regions in the left inferior parietal lobe support phonological short-term memory: evidence for a buffer account? *Cereb Cortex* 29:1398–1413.