# Dissociated neural representations of content and ordinal structure in auditory sequence memory

Ying Fan[1-3], Qiming Han[2-4], Simeng Guo[5], Huan Luo[1-3,6,*]

[1] School of Psychological and Cognitive Sciences, Peking University

[2] PKU-IDG/McGovern Institute for Brain Research, Peking University

[3] Beijing Key Laboratory of Behavior and Mental Health, Peking University

[4] Peking-Tsinghua Center for Life Sciences, Peking University

[5] Yuanpei College, Peking University, Beijing 100871, China

[6] Lead contact

**Running title:** factorized structure-content coding in auditory memory

**Address for Correspondence:**

* Huan Luo

School of Psychological and Cognitive Sciences, Peking University

PKU-IDG/McGovern Institute for Brain Science, Peking University

Room 1703, 52 Haidian Road, Beijing, 100087, China

huan.luo@pku.edu.cn

**Abstract**

When retaining a sequence of auditory tones in working memory (WM), two forms of information – frequency (content) and ordinal position (structure) – have to be maintained in the brain. Here, we employed a time-resolved multivariate decoding analysis on content and structure information separately to examine their neural representations in human auditory WM. We demonstrate that content and structure are stored in a dissociated manner and show distinct characteristics. First, each tone is associated with two separate codes in parallel, characterizing its frequency and ordinal position, respectively. Second, during retention, a structural retrocue reactivates structure but not content, whereas a following white noise triggers content but not structure. Third, structure representation remains unchanged whereas content undergoes a transformation throughout memory progress. Finally, content reactivations during retention correlate with WM behavior. Overall, our results support a factorized content-structure representation in auditory WM, which might help efficient memory formation and storage by generalizing stable structure to new auditory inputs.

**Key words**: auditory working memory, sequence structure, factorization, multivariate decoding, reactivations, hidden state

## Introduction

Memories are not stored in fragments; instead, multiple items or events are constantly linked and organized with each other according to certain relationship. 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, and rehearsing a piece of favorite melody requires sorting of each musical tone in certain temporal order. Thus, 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., the ordinal position for each tone) – in brain activities.

Although content representation in WM has been widely studied (e.g., Kumar et al., 2016; Lundqvist et al., 2016; Murray et al., 2017; Huang et al., 2018; Kamiński et al., 2020; Quentin et al., 2019; Sutterer et al., 2019; Wolff et al., 2020), how the brain encodes and stores the somewhat abstract structure that presumably organizes and sorts fragmented contents remains largely unknown, not to mention their coexisting representations in WM. Previous studies demonstrate brain regions for structure information (Baldassano et al., 2017; Carpenter et al., 2018; DuBrow & Davachi, 2016; Jenkins & Ranganath, 2010; Kalm & Norris, 2014), as well as its neural oscillatory correlates (Hsieh et al., 2011; Hsieh & Ranganath, 2014; Liu et al., 2020; Roberts et al., 2013). About their relationship in WM, content and structure could be stored in a combinational way, i.e., each item is indexed by a specific content-structure combination (Kikumoto & Mayr, 2020; Komorowski et al., 2009; Naya et al., 2017), or in a factorized manner, i.e., content and structure for a given item are encoded independent of each other (Bengio et al., 2013; Higgins et al., 2016; Kalm et al., 2014; Liu et al., 2019; Marshuetz, 2005). It is posited that factorized coding could help efficient memory formation by enabling rule generalization across different inputs (Behrens et al., 2018; Friston & Buzsáki, 2016).

The present study aimed to examine how frequency content and ordinal structure of an auditory tone sequence are encoded and maintained in human WM. To this end, we used a time-resolved multivariate decoding analysis on the electroencephalography (EEG) signals recorded from human participants in two auditory sequence memory experiments. Importantly, the decoding analysis of frequency content and ordinal structure were performed independent of each other, by marginalizing the other one, to particularly test the content-structure factorization idea. In other words, if content and structure are indeed represented in a completely combinational format (i.e., specific codes for a particular combination), the current analysis would expect neither content nor structure representations.

Furthermore, since the EEG-recorded macroscopic brain activities during the delay period tend to stay in an activity-silent state (e.g., Mongillo el al., 2008; Rose et al., 2016; Sprague et al., 2016; Stokes, 2015) or lack enough sensitivity to reflect the sustained neuronal firing (e.g, Funahashi et al., 1989; Masse, 2017; Mendoza-Halliday et al., 2014; Zhou et al., 2013), it is difficult to directly and reliably assess the stored WM representations. We thus employed an impulse-response approach (Wolff et al., 2017) to transiently perturb the neural network and measure the elicited WM information presumably stored in the network. Specifically, two trigger events were presented successively during retention – a structure retrocue and a white-noise auditory impulse. The former, as an abstract structure cue, would presumably reactivate the stored structure representation (i.e., ordinal position), while the latter is hypothesized to trigger content information (i.e., tone frequency) by perturbing the auditory cortex where tone contents likely reside. Notably, the factorization view does not assume that the retrocue could only trigger structure information while the white noise only reactivates contents.

Our results demonstrate that content and ordinal structure of an auditory tone sequence are encoded in a factorized manner and show distinct characteristics in auditory WM. First, each item in the list is associated with two separate codes, characterizing its specific frequency and ordinal position in the list, respectively.

Second, a structure retrocue during the delay period activates structure representation but not content, whereas an auditory white noise triggers content-related response but not structure, implying they might be stored in different brain regions. Third, structure representation remains largely unchanged whereas content coding is dynamically transformed throughout memory progress. Finally, content representations that are reactivated during retention correlate with subsequent auditory WM performance in each participant. Taken together, our results support a dissociated form of content and structure representation in auditory WM. This factorized mechanism might facilitate memory formation through efficient information organization and rapid structure implementation on new contents.

## Results

### Experimental procedure and factorization-oriented decoding analysis

In Experiment 1, thirty human participants performed an auditory delayed-match-to-sample working memory task while their 64-channel electroencephalography (EEG) activities were recorded. Each participant performed 1080 trials in total (6 hours in two separate days). As shown in Figure 1A, in each trial, three pure tones with different frequencies that were pseudo-randomly selected from six fixed values ($f1: 381\ Hz$, $f2: 538\ Hz$, $f3: 762\ Hz$, $f4: 1077\ Hz$, $f5: 1524\ Hz$, $f6:\ 2155\ Hz$) 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 recalled and tested at the end of the trial. Two seconds after the retrocue, a 100-ms white-noise auditory impulse was presented. Finally, 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. Note that the frequency of the target tone was determined in a pretest so that the overall memory performance would be around 75%. Indeed, the behavioral

performance for the auditory WM task was 75.35% (SE = 1.00%).

A time-resolved representational similarity analysis (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 Methods). Notably, instead of binary decoding, the RSA analysis here is based on a hypothesis that the neural representational similarity is proportional to the factor-of-interest similarity, i.e., the neural response for $f1$ $(381\ Hz)$ and $f2$ $(538\ Hz)$ would be less dissimilar than that for $f1$ $(318\ Hz)$ and $f5$ $(1524\ Hz)$. To quantify the representational strength, we next calculated the linear regression ($\beta$) between the factor dissimilarity and the corresponding neural representational dissimilarity for frequency and ordinal position, respectively. Moreover, it is noteworthy that the decoding analysis employed here specifically aims to test the content-structure factorization representation, i.e., decoding one factor (e.g., frequency) while marginalizing the other factor (e.g., ordinal position). In other words, if content and structure are completely encoded in a combinational manner, we would not observe the content decoding, since the content codes would vary with its specific linked position. Finally, All the analyses expect that for correct - wrong trial comparisons were conducted on behaviorally correct trials only.

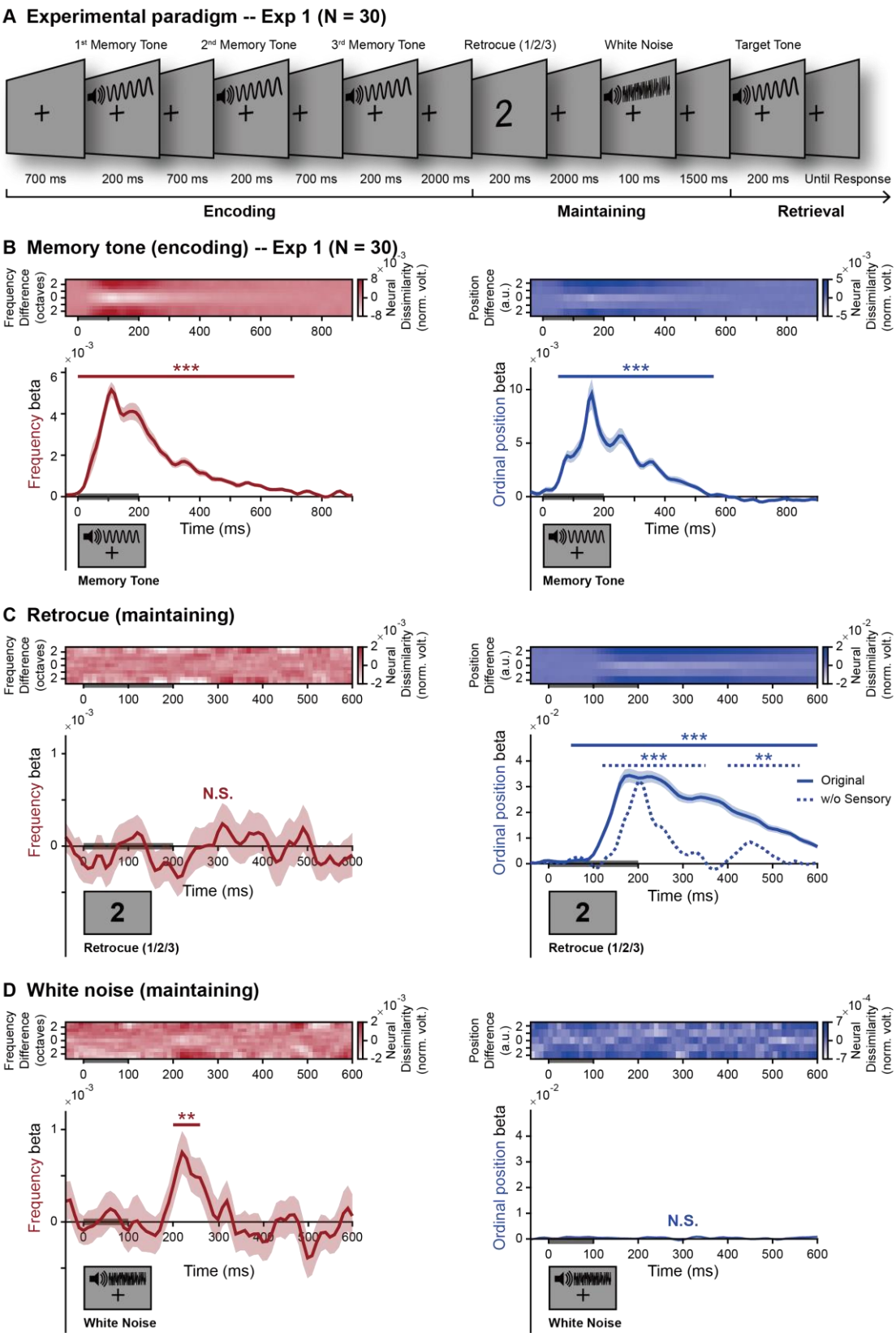


**Figure 1. Experimental paradigm and neural representations of content and structure during encoding and retention periods (Exp 1)**

(A) Experimental 1 paradigm. In each trial, three pure tones (1st, 2nd, and 3rd Memory

Tones) with different frequencies (selected from 381 Hz, 538 Hz, 762 Hz, 1077 Hz, 1524 Hz 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' or '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 to 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. Upper 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. Lower panels: Grand average (N = 30, mean ± SEM) beta 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 indicates memory tone presentation. (C) Same as B but during maintain period after retrocue. Gray bar on x-axis indicates retrocue presentation. Dotted blue line (right panel): the sensory-removed decoding strength for ordinal position; corrected. (D) Same as B but during maintaining period after white-noise auditory impulse. Gray bar on x-axis indicates white-noise presentation. Note that only correct trials were analyzed in each participant. (Shaded area represents ± 1 SEM across participants. ***: $p < 0.001$; **: $p < 0.05$; solid line: corrected using cluster-based permutation test, cluster-forming threshold $p < 0.05$).

**Dissociated structure and content representation during WM encoding**

During the encoding period, each of the three tones could be characterized by two factors, i.e., frequency ($f1: 381\ Hz,\ f2: 538\ Hz,\ f3: 762\ Hz,\ f4: 1077\ Hz,\ f5: 1524\ Hz,\ f6:\ 2155\ Hz$) and ordinal position (1st, 2nd, 3rd). Therefore, we could assess the neural representations of content and structure, independently, from the same EEG response for each tone. Specifically, when examining frequency encoding, each tone would be labelled in terms of its frequency, regardless of its ordinal position, whereas when analyzing structure representation, we would tag tones based on their ordinal positions regardless of the frequencies. Moreover, a cross-validated confound regression approach (Snoek et al., 2019) was employed to regress out possible influences from global field power courses for the three tones on the multivariate decoding analysis.

The upper panels of Figure 1B plot the time-resolved neural representational dissimilarity as a function of factor 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 response contains information for both content and structure. The lower panel of Figure 1B shows the corresponding regression weights of the dissimilarity matrices, for frequency ($\beta_{freq}$, left) and ordinal position ($\beta_{pos}$, right), separately. Specifically, both the frequency representation $\beta_{freq}$ (left; $p < 0.001$, corrected) and the ordinal position representation $\beta_{pos}$ (right; $p < 0.001$, corrected) emerged shortly after the onset of each tone.

In summary, during the encoding period when a sequence of tones is presented to be maintained in auditory WM, each sound is signified by two separate neural codes that characterize its content (frequency) and structure (ordinal position), respectively (see Figure S1 for another analysis to confirm the factorized encoding of frequency and ordinal position).

**Dissociated structure and content representation during WM retention**

After establishing the separate representations of frequency and ordinal position during the encoding phase, we next examined their maintenance during the delay period. To this end, the RSA analysis would be performed in reference to the memory-related information, e.g., a trial with retrocue of '2' and $f4$ $(1077\ Hz)$ as the 2nd memorized tone's frequency would be labelled as '2' for ordinal position and $f4$ for frequency. Moreover, the whole delay period was divided into two phases – after the retrocue (Figure 1C) and after the white-noise impulse (Figure 1D), for illustration purpose.

First, the brain activities remained in an 'activity-silent' state prior to either retrocue or white noise, in terms of both global field power and decoding performance (see Figure S2). Interestingly, the memory-related structure and content information

were successfully reactivated by triggering events during retention and displayed largely dissociated activation profiles (Figure 1CD). As shown in Figure 1C, the ordinal position representation (right, blue line) emerged right after the positional retrocue ($p < 0.001$, corrected), whereas the tone frequency information (left, red line) failed to be activated ($p > 0.5$, corrected). In contrast, as shown in Figure 1D, the white-noise auditory impulse successfully activated the tone frequency representation (left, red line; 200 – 260 ms: $p = 0.017$, corrected) but not the structure representation (right, blue line; $p > 0.359$, corrected). Therefore, contents and ordinal structure tend to be maintained in WM independent of each other (i.e., successful decoding using this analysis) and could be efficiently reactivated by white noise and retrocue, respectively. Meanwhile, it is noteworthy that the factorization view does not necessitate that the retrocue could only trigger structure information while the white noise only reactivates contents (see Discussion).

Importantly, since the retrocue stimuli are different (e.g., '1', '2', '3'), the structure representation after the retrocue (Figure 1C, 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 (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 1C, the sensory-removed decoding performance (right, dotted blue line), although largely dropped compared to the original results (solid line), was still significant over the retention period, 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 that would influence the structure representations, an issue we further addressed using an encoding-to-maintaining generalization analysis (see Figure 2).

Furthermore, we examined whether the presented white noise during retention

would interfere with auditory WM. A control behavioral experiment with completely different participants (N =19; see Figure S3) revealed that the auditory sequence WM behavior with or without noise presentation during retention showed no difference in both accuracy and reaction time. This confirms that the presented white noise serves as a 'neutral' probe to assess the information retained in WM network and would not influence the WM behavior.

**Distinct cross-temporal generalization for structure and content**

Having confirmed the neural representations during both encoding (Figure 1A) and maintaining period for content (Figure 1C, after white noise) and structure (Figure 1D, after retrocue), we next conducted an encoding-to-maintaining generalization analysis to estimate whether these two phases share similar coding formats for frequency and ordinal position, respectively. This is also motivated by recent findings revealing dynamic (i.e., non-generalizable) representations for contents (e.g., Barak et al., 2010; Lundqvist et al., 2018; Rademaker et al., 2019; Pastalkova et al., 2008; Trübutschek et al., 2017; Wolff et al., 2017) and constant neural codes (i.e., generalizable) for abstract concepts (Bueti & Walsh, 2009; Luyckx et al., 2019; Parkinson et al., 2014). Put another way, if the neural codes (i.e., frequency, ordinal position) for encoding (Figure 1B) could be successfully generalized to those elicited by the corresponding trigger events (i.e., white noise, retrocue) during retention (Figure 1CD), this would support a stable WM coding, otherwise a dynamic transformation.

We first used the same RSA approach as performed before to examine the cross-temporal generalization, except that here the response dissimilarity was calculated across the encoding and retention activations. As shown in Figure 2B, only the structural information showed a significant cross-temporal generalization (right; 230 – 280 ms: $p = 0.026$, corrected), whereas the content information did not (left; $p > 0.089$, corrected). To further verify these results, we also used a lasso-regularized logistic regression models to test the cross-temporal generalization. Notably, instead

of a dissimilarity-based analysis in RSA, this approach is based on binary classifications. As shown in Figure 2C, likewise, only the structure information showed significant encoding-to-maintenance generalization (right; 90 – 570 ms: $p < 0.001$, corrected), whereas frequency information did not (left; $p > 0.438$, corrected).

Taken together, structure and content are endowed with distinct representational transformation properties, i.e., structure representation remains relatively stable from encoding to maintaining periods, while the content representation undergoes a dynamic change. Importantly, the successful cross-temporal generalization results for ordinal position also provide convincing evidence for the genuine structure representations reactivated by the retrocue, since there was no cues (i.e., '1', '2', '3' stimuli) at all presented during the encoding period yet the structural codes could still be generalized to that during retention.

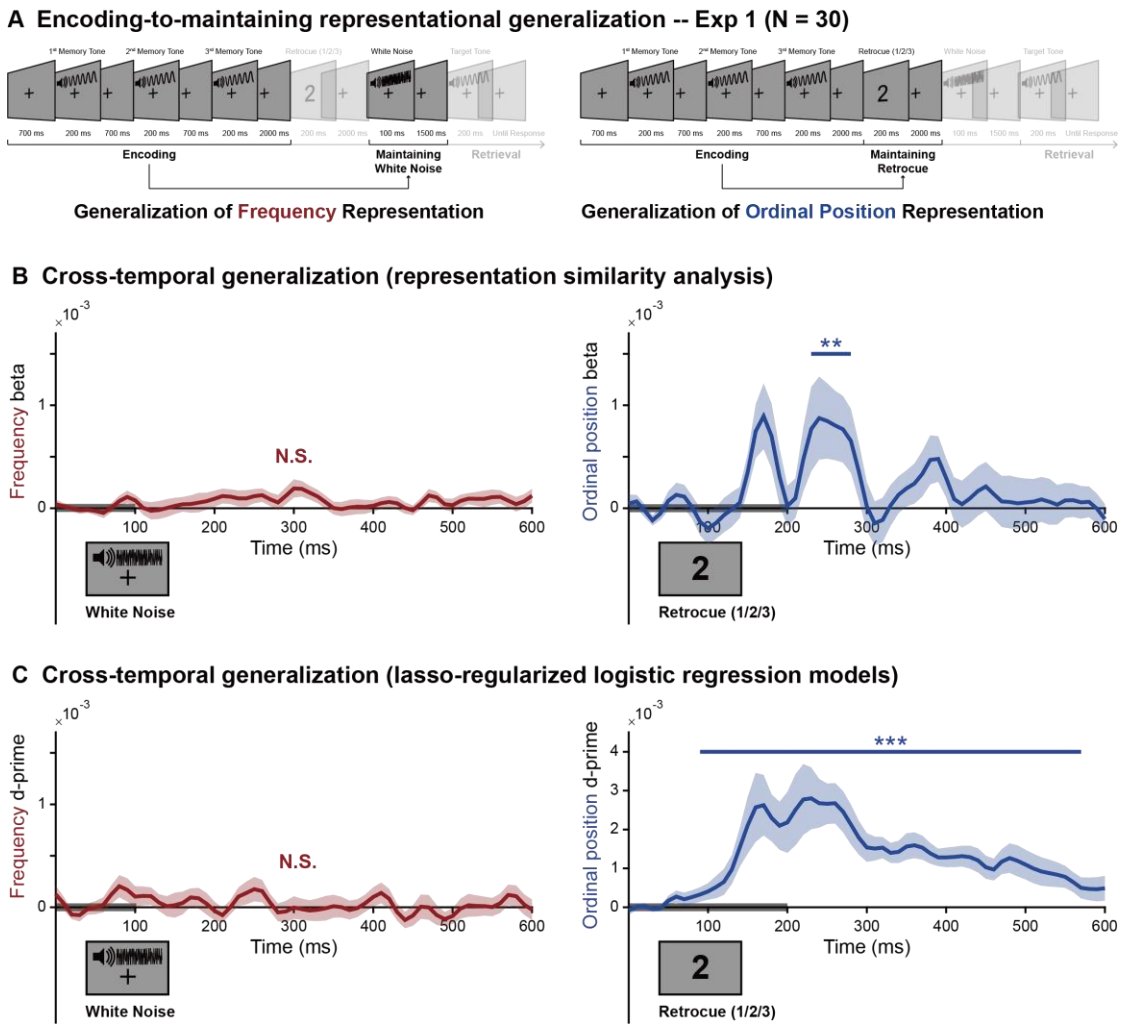

**Figure 2. Encoding-to-Maintaining representational generalization (Exp 1)** (A) Illustration of the cross-temporal generalization from encoding to maintaining periods for frequency (left) and ordinal position (right). Left: frequency generalization, i.e., encoding-phase response (within 90 - 130 ms) were used as train data to test the maintaining-phase responses at each time point after white-noise auditory impulse. Right: ordinal position generalization, i.e., encoding-phase response (within 140 - 180 ms) were used as train data to test the maintaining-phase responses at each time point after retrocue. (B) Grand average (N = 30, mean ± 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 indicates white-noise (left) and retrocue (right). (C) Same as B but using lasso-regularized logistic regression models. Note that only correct trials were analyzed in each participant. (Shaded area represents ± 1 SEM across participants. ***: $p < 0.001$; **: $p < 0.05$; solid line: corrected using cluster-based permutation test, cluster-forming threshold $p < 0.05$;).

**Content reactivations correlate with memory behavior**

Finally, we evaluated whether the neural representations of content and structure retained in WM, which are reactivated by the retrocue and white noise, have any behavioral consequence, by performing the same RSA analysis on correct trials and incorrect trials separately, in each participant. We ensured equal numbers of correct and incorrect trials to make a fair comparison (see details in Methods). Interestingly, as shown in Figure 3A, 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 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), which also displayed a trend of behavioral relevance (wrong vs. correct: 440 – 450 ms: $p = 0.351$, corrected; maximum $p = 0.016$, uncorrected).

Meanwhile, the structure reactivations (i.e., encoding-to-maintaining generalization, given its genuine indexing of structure information independent of

retrocue stimuli) displayed no behavioral relevance. As shown in Figure 3B and 3C, comparable ordinal position representations between correct and wrong trials during retention were observed using both RSA (Figure 3B, $p > 0.138$, corrected) and lasso-regularized logistic regression model (Figure 3C, $p > 0.106$, corrected).

Finally, as a control analysis, we examined the behavioral relevance of neural response during the encoding period by comparing the content (Figure 3D) and structure (Figure 3E) representations between correct and wrong trials. As shown in Figure 3DE, behaviorally-different trials displayed similar decoding performance for both frequency ($p > 0.287$, corrected) and ordinal position ($p > 0.5$, corrected), suggesting that the deteriorating WM performance in wrong trials were not attributable to encoding failure.

In sum, frequency information maintained in auditory WM is associated with subsequent recalling performance. The better the frequency information are retained in WM, which could be probed and assessed by the impulse-response approach during retention, the better the subsequent memory performance.

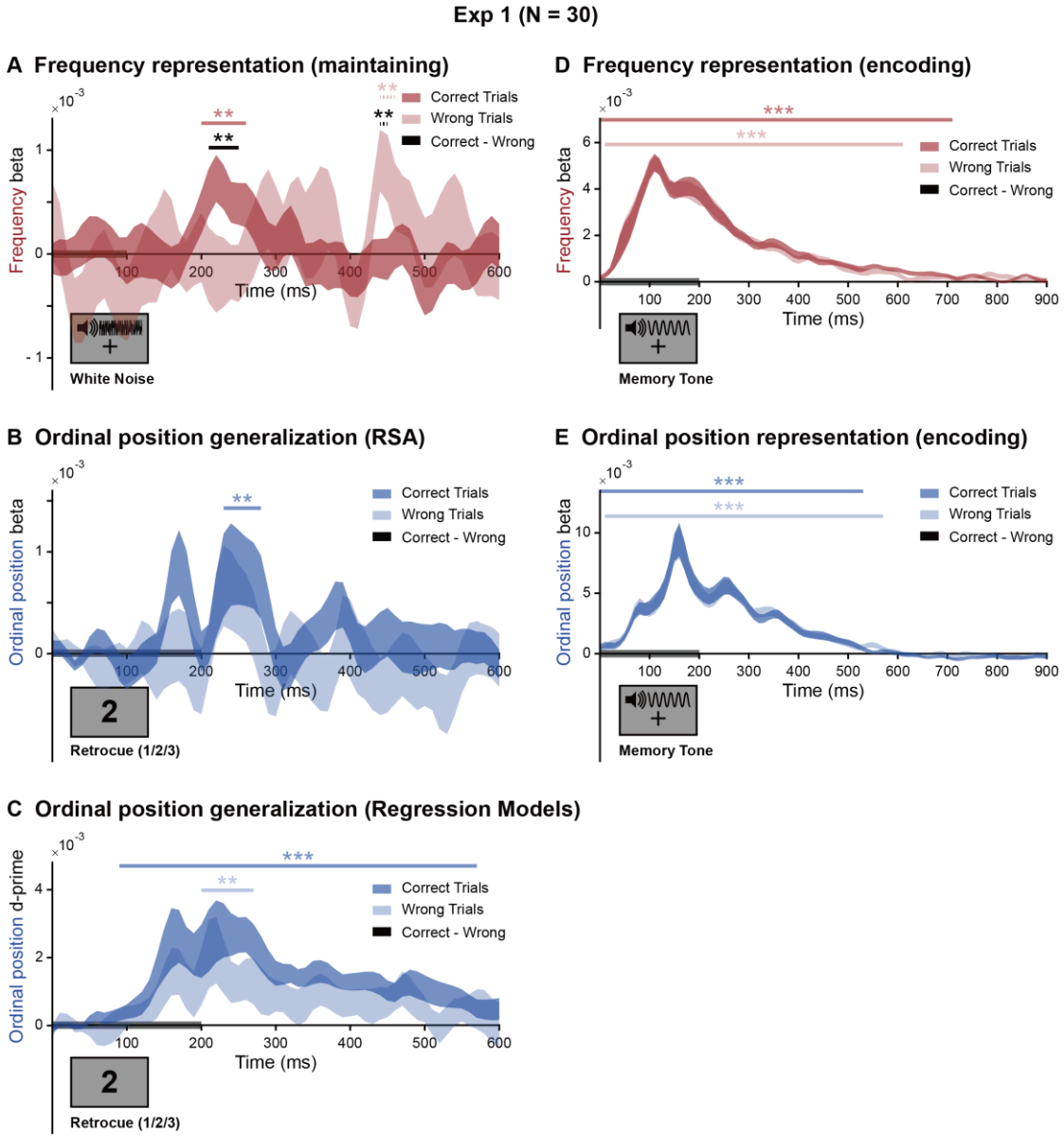


**Figure 3. Behavioral correlates of content and structure representations in auditory WM (Exp 1)**

(A) Grand average (N = 30, mean ± 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 indicates white-noise presentation. (B) Grand average (N = 30, mean ± 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 indicates retrocue presentation. (C) Same as B but using lasso-regularized logistic regression models. (D) Same as A but during encoding period after auditory tones. Gray bar on x-axis indicates tone presentation. (D) Same as B but during encoding period after auditory tones. Gray bar on x-axis indicates tone presentation. Note that 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:

corrected using cluster-based permutation test, cluster-forming threshold $p < 0.05$; dotted line: uncorrected).

**Experiment 2**

In Experiment 1 (Figure 1A), participants were instructed to retain a list of auditory tones in WM and recall the tone frequency according to the retrocue (1st, 2nd, 3rd tone). Meanwhile, since the recalling task only tested the tone frequency (i.e., content information), participants might have discarded the structure information (i.e., the ordinal position) after the retrocue. This would constitute an alternative interpretation for why the white-noise impulse failed to trigger structure information (Figure 1CD). To address this possibility, we performed Experiment 2 (N = 18), during which the participants did exactly the same auditory sequence memory task as Experiment 1 but needed to additionally recall the position of the memorized tone (Figure S4A). In other words, Experiment 1 and Experiment 2 shared the same experiment design except the latter required participants to explicitly hold both the ordinal position and frequency information in auditory WM.

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 4, Experiment 2 largely replicated the results of Experiment 1 (see Figure S4B for the results during encoding period in Exp 2). First, the retrocue elicited structure representation (Figure 4A, right; 60 – 600 ms: $p = 0.003$, corrected) but not frequency (Figure 4A, left; $p > 0.337$, corrected), while the white noise reactivated content (Figure 4B, left, dark red; 250 – 320 ms: $p = 0.006$, corrected) but still failed to drive structure information (Figure 4B, right, $p > 0.350$, corrected), consistent with Experiment 1. Moreover, we replicated the cross-temporal generalization results, i.e., a stable encoding-to-maintaining generalization for structure (Figure 4C, right, dark blue; 140 – 180 ms: $p = 0.046$, corrected) but not for content (Figure 4C, left, $p > 0.251$, corrected). Experiment 2 also showed the behavioral correlates of reactivation profiles, i.e., the correct trials showed better frequency representation after the white noise than

incorrect trials (Figure 4B, left, black star, 250 – 280 ms: $p = 0.054$, corrected). Meanwhile, same as Experiment 1, correct and wrong trials elicited comparable structure reactivations (i.e., encoding-to-maintaining generalization, given its genuine indexing of structure information independent of retrocue stimuli) (Figure 4C, right $p > 0.158$, corrected). Note that we did not divide trials based on ordinal position performance given the high accuracy here (98.80%±0.22%).

Finally, we combined Experiment 1 and Experiment 2 (N = 48) to summarize the behavioral correlates of the reactivations, since they shared the same design (through encoding and retention periods) and showed similar response patterns. The white noise triggered stronger content reactivation for correct than incorrect trials (Figure 4D, black, 210 – 270 ms: $p = 0.011$, corrected), while the retrocue elicited similar structure reactivations (i.e., cross-temporal generalization) for correct versus incorrect trials (Figure 4E, $p > 0.148$, 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 of correct trials (430 – 450 ms: $p = 0.178$, corrected; maximum $p = 0.045$, uncorrected). Taken together, content representations during retention is associated with subsequent corresponding WM behavioral performance.

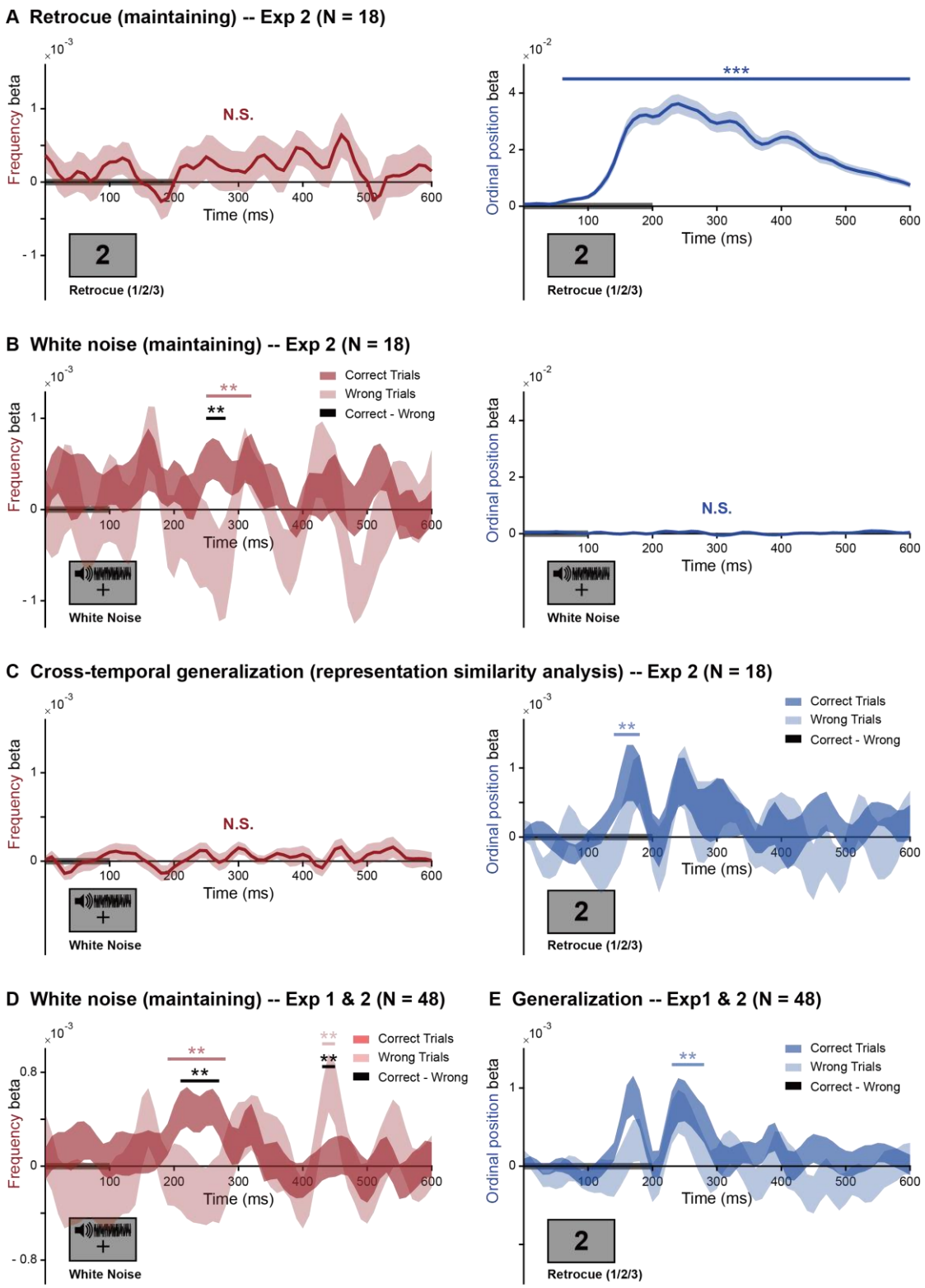


**Figure 4. Experiment 2 results**

(A) Neural representations of content (i.e., frequency) and structure (i.e., ordinal position) during maintaining period after retrocue. Grand average (N = 18, mean ± SEM) beta 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. (B) Left: Grand average (N = 18, mean ± SEM) frequency decoding performance after white noise, for correct (dark red) and wrong (light red) trials. Right: Grand average ordinal position decoding performance after white noise. (C) Grand average (N = 18, mean ± SEM) encoding-to-maintaining generalization results for frequency (left) and ordinal position (right) for correct (dark blue) and wrong (light blue) trials. (D) Grand average (N = 48, mean ± SEM) frequency decoding performance after white noise, for correct (dark red) and wrong (light red) trials, by combining Exp 1 and Exp 2. (E) Grand average (N = 48, mean ± SEM) encoding-to-maintaining generalization results for ordinal position after retrocue, for correct (dark blue) and wrong (light blue) trials, by combining Exp 1 and Exp 2. Note that the analyses were performed on correct and wrong trials with equal numbers in each participant. (Shaded area represents ± 1 SEM across participants. ***: $p < 0.001$; **: $p < 0.05$; solid line: corrected using cluster-based permutation test, cluster-forming threshold $p < 0.05$; dotted line: uncorrected).

## 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 dissociated manner in human auditory WM. Each item is signified by two separate neural codes – content and structure – which could be reactivated later during retention, by retrocue and white-noise impulse, respectively. Furthermore, they demonstrate distinct temporal stability throughout memory process, i.e., stable representation for structure and dynamical coding for content, and the reactivations of content contributes to subsequent WM recognition performance. Overall, the human brain extracts and separates structure and content information from auditory inputs and maintains them in different formats. This factorized structure-content scheme might benefit WM via implementing abstract, stable structure on new content inputs.

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 essentially influences perception (i.e., Gestalt principle; Wagemans et al., 2012), memory (Davachi & DuBrow, 2015; DuBrow & Davachi, 2013; Gershman et al., 2013) 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 encoding of spatial (Eichenbaum et al., 1999; Muller, 1996; O'Keefe & Dostrovsky, 1971; O'keefe & Nadel, 1978) or temporal structure (Eichenbaum, 2014; MacDonald et al., 2011; Manns et al., 2007; Paz et al., 2010). Interestingly, hippocampus might play a more general function in representing relationships among events even in nonspatial tasks, i.e., cognitive map (Aronov et al., 2017; Garvert et al., 2017; Schuck & Niv, 2019). In addition, parietal regions are also found to be involved in structural representation and learning (Bueti et al., 2009; Feigenson et al., 2004; Parkinson et al., 2014; Summerfield et al., 2020; Walsh, 2003). Our study focused on ordinal information of a tone sequence, a typical and important

type of relational structure to sort contents in auditory experience. Neural coding of ordinal position has been found in both animal recordings (Fortin et al., 2002; Manns et al., 2007; Naya & Suzuki, 2011) and neuroimaging studies (Carpenter et al., 2018; DuBrow et al., 2016; Hsieh et al., 2011, 2014; Jenkins et al., 2010; Kalm et al., 2014; Liu et al., 2020; Rajji et al., 2017; Roberts et al., 2013). A recent study revealed that even spontaneous brain activities contain temporal order information in an abstract generalization task (Liu et al., 2019). Our results are thus consistent with previous findings and further expand to auditory WM.

Different from visual studies, only a few studies have directly assessed the neural correlates of auditory WM (Albouy et al., 2017; Luo et al., 2013; Weisz et al., 2020; Wolff et al., 2020). Here, by combining RSA-based decoding and impulse-response approach, we demonstrate that during retention, a white-noise impulse that does not contain any memory-related information could successfully activate the tone frequency information held in auditory WM. This is consistent with the hidden-state WM view, which posits that information could be stored in synaptic weights via short-term neural plasticity principles (Lewis-Peacock et al., 2012; Masse et al., 2020; 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 the WM network and would not modify WM contents (also see the control experiment, Figure S3). On the other hand, not all the triggering events could reactivate the stored contents, i.e., the frequency information seemingly could not be triggered by the retrocue and the ordinal position could not be activated by white noise, 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, while structure information might be maintained in high-level areas, e.g., parietal region (Bueti et al., 2009; Parkinson et al., 2014; Summerfield et al.,

2020; Walsh, 2003).

The decoding method employed here is factorization-oriented, i.e., decoding the factor of interest while marginalizing the other. Therefore, even if the retrocue elicits content representation or the white noise reactivates structure information during retention, the factorization view still holds. In other words, the factorization view does not necessarily entail that distinct trigger events reactive different types of information in a dissociated manner. For example, concurrent representations of content and structure during the encoding period (Figure 1B) as observed here still support the content-structure dissociated codes when auditory tones are physically presented. Future studies need to be performed to examine whether different task contexts or other triggering events could reactivate both content and structure information during retention. Furthermore, our results do not exclude the presence of combinational coding, as previously revealed in animal recordings (Chen and Naya, 2020; Hsieh et al., 2014; Naya et al., 2011; Tsao et al., 2013; Tsao et al., 2018). Due to the limited sensitivity of EEG recordings, the factorized codes might be more robust to be assessed here.

The white-noise-elicited tone frequency representation correlates with subsequent recognition performance, further supporting its genuine indexing of content maintenance in auditory WM. Interestingly, a trend of late content reactivation for incorrect trials (Figure 3A, light red; Figure 4D, light red) 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, i.e., 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). WM computational modelling that incorporates competition between items also predicts late reactivation for weakly stored WM items (Mongillo et al., 2008). Therefore, besides reactivation strength, the latency of content reactivations might serve as another index to assess subsequent WM performance.

Our results reveal a dissociated representation of content and structure in auditory WM. What is the benefit of the factorized representation? An apparent advantage is that it would allow rapid and versatile transfer of the structure information to new contents (Behrens et al., 2018; Friston et al., 2016; Kaplan et al., 2017; Luyckx et al., 2019; Tse et al., 2007; 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; Cowan et al., 2013; Fougnie & Alvarez, 2011). 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 representation for structure and dynamic transformation for content. The transformed representation for WM contents has been widely found in previous studies using a wide variety of experimental paradigms (e.g., Barak et al., 2010; Meyers et al., 2008; Spaak et al., 2017; Yue et al., 2019; with cue: Sprague et al., 2016; Wolff et al., 2017, 2020; Panichello et al., 2020; Parthasarathy et al., 2017, 2019; Yu et al., 2020). Our results are thus fully consistent with previous findings. One interpretation for the dynamic transformation is that the WM content is coded in a way that is optimized for subsequent behavioral demands (Panichello & Buschman, 2020; Myers et al., 2017). Another is that the content information is transformed into a subspace to resist distractions from new inputs (Libby & Buschman, 2019; Murray et al., 2017). In contrast, the ordinal structure displays a stable representation over memory course, also in line with previous findings (Luyckx et al., 2019; Walsh, 2003) and the factorization idea, i.e., stable structure coding would be advantageous for its generalization role in memory formation.

Taken together, content (tone frequency) and structure (ordinal position) are the two basic formats of information to be maintained while retaining a tone sequence. Our findings demonstrate that they are represented and stored in a largely dissociated

way and display distinct dynamics, which echo their varied functions in memory formation – detailed and dynamic contents vs. abstract and stable structure.

**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.

**Conflicts of interest**

The authors declare no conflicts of interest.

**Acknowledgments**

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

## METHODS

### Participants

Thirty (17 females, mean age 23.4, range 19-27 years) and twenty (8 females, mean age 19.7, range 18–25 years) healthy participants with normal or corrected-to-normal vision were recruited in Experiment 1 and Experiment 2, respectively, after providing written informed consent. Two participants in Experiment 2 were excluded due to 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 (MathWorks) and Psychophysics Toolbox (Psychtoolbox-3; Brainard, 1997). The visual stimuli were presented on a Display ++ LCD screen with a resolution of 1920 by 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 approximately 65 dB (62.1 to 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 (Figure 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 Hz, 538 Hz, 762 Hz, 1077 Hz, 1524 Hz and 2155 Hz) were presented sequentially. The duration of each memory tone was 200 ms, with inter-stimulus interval of 700 ms. 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 judgement 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. 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 two separate days, for each participant. Each session lasted approximately 3 hours.

**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 Figure S4A). Furthermore, to control the temporal length of the experiment, there were other timing parameter adjustments, i.e., the interval between the 3rd 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, in order to prevent motor preparation during the maintaining period, the correspondence between particular ordinal positions and reaction keys are randomly set from six combinations in each trail. Same as Experiment 1, there were 1080 trials in total, which were separated into two sessions being completed in two separate days, for each participant. Each session lasted approximately 3 hours.

**EEG acquisition and pre-processing**

The EEG signals were acquired using a 64-channel actiCAP (Brain Products) and two BrainAmp amplifiers (Brain Products). The data was 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 below 10 kΩ. The EEG data was referenced to the average value across all channels, down-sampled 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. Epochs with extremely high noises by visual inspection were manually excluded from the following analyses.

**Time-resolved multivariate decoding**

Note that only the correct trials in each participant were used for further analysis, except the behavioral correlates analyses (Figure 3, Figure 4) during which the same number of correct and wrong trials in each participant were analyzed and compared (see details in later Methods).

First, a time-resolved representational similarity analysis (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; Grootswagers 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 employed 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. All the subsequent decoding analysis were conducted on the corrected data.

The RSA analysis 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*2 = 128 in total) for a further 8-fold cross-validation decoding approach. The Mahalanabis 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 & 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 beta 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), were then averaged across the 50 partitions as the final decoding accuracy, at each time point and for each participant. Note that the RSA analysis was performed for frequency and ordinal position, independently, using the same EEG data but with different feature dimensions.

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 (dotted line in Figure 1C, right).

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)-folds correct trials were used as training data, and the left 1-fold 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). Note that as shown in Figure 1CD, 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 employed 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 Mahalanibos distance was computed between response 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 $\beta$ representing weights for each feature and $\alpha$ 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-prime values, which represent the cross-temporal generalization decoding performance.

For both methods, the decoding results were further baseline corrected (0 to 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**

Non-parametric sign-permutation test (Maris & Oostenveld, 2007; customized analysis codes) 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.

## Reference


Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron, 94*(1), 193-206. e195.

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature, 543*(7647), 719-722.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron, 95*(3), 709-721. e705.

Barak, O., Tsodyks, M., & Romo, R. (2010). Neuronal population coding of parametric working memory. *Journal of Neuroscience, 30*(28), 9424-9430.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, 49(6), 1622-1631.Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron, 100*(2), 490-509.

Bengio, Y., Mesnil, G., Dauphin, Y., & Rifai, S. (2013). *Better mixing via deep representations.* Paper presented at the International conference on machine learning.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*(4), 433-436.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1525), 1831-1840.

Carpenter, A. F., Baud-Bovy, G., Georgopoulos, A. P., & Pellizzer, G. (2018). Encoding of serial order in working memory: neuronal activity in motor, premotor, and prefrontal cortex during a memory scanning task. *Journal of Neuroscience, 38*(21), 4912-4933.

Chen, H., & Naya, Y. (2020). Forward processing of object–location association from the ventral stream to medial temporal lobe in nonhuman primates. *Cerebral Cortex*, 30(3), 1260-1271.

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 731.

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: the role of prediction and context. *Trends in cognitive sciences, 19*(2), 92-99.

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General, 142*(4), 1277.

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of learning and memory, 134*, 107-114.

Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience, 15*(11), 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*(2), 209-226.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences, 8*(7), 307-314.

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience, 5*(5), 458-462.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of vision*, 11(12), 3-3.

Friston, K., & Buzsáki, G. (2016). The functional anatomy of time: what and when in the brain. *Trends*

*in cognitive sciences, 20*(7), 500-511.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology*, 61(2), 331-349.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *Elife, 6*, e17086.

Gershman, S. J., Schapiro, A. C., Hupbach, A., & Norman, K. A. (2013). Neural context reinstatement predicts memory misattribution. *Journal of Neuroscience, 33*(20), 8590-8595.

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience, 29*(4), 677-697.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., . . . Lerchner, A. (2016). beta-vae: Learning basic visual concepts with a constrained variational framework.

Hsieh, L.-T., Ekstrom, A. D., & Ranganath, C. (2011). Neural oscillations associated with item and temporal order maintenance in working memory. *Journal of Neuroscience, 31*(30), 10803-10810.

Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron, 81*(5), 1165-1178.

Huang, Q., Jia, J., Han, Q., & Luo, H. (2018). Fast-backward replay of sequentially memorized items in humans. *Elife, 7*, e35164.

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *Journal of Neuroscience, 30*(46), 15558-15565.

Kalm, K., & Norris, D. (2014). The representation of order information in auditory-verbal short-term memory. *Journal of Neuroscience, 34*(20), 6879-6886.

Kamiński, J., Brzezicka, A., Mamelak, A. N., & Rutishauser, U. (2020). Combined Phase-Rate Coding by Persistently Active Neurons as a Mechanism for Maintaining Multiple Items in Working Memory in Humans. *Neuron*.

Kaplan, R., Schuck, N. W., & Doeller, C. F. (2017). The role of mental maps in decision-making. *Trends in Neurosciences, 40*(5), 256-259.

Kikumoto, A., & Mayr, U. (2020). Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. *Proceedings of the National Academy of Sciences, 117*(19), 10603-10608.

Komorowski, R. W., Manns, J. R., & Eichenbaum, H. (2009). Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. *Journal of Neuroscience, 29*(31), 9918-9929.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*, 4.

Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., & Griffiths, T. D. (2016). A brain system for auditory working memory. *Journal of Neuroscience, 36*(16), 4492-4505.

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management, 30*(4), 110-119.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24*(1), 61-79.

Libby, A., & Buschman, T. J. (2019). Rotational Dynamics Reduce Interference Between Sensory and

Memory Representations. *bioRxiv*, 641159.

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. (2019). Human replay spontaneously reorganizes experience. *Cell, 178*(3), 640-652. e614.

Liu, X., Ranganath, C., Hsieh, L., Hurtado, M., Niendam, T., Lesh, T., . . . Ragland, J. (2020). Task-specific Disruptions in Theta Oscillations during Working Memory for Temporal Order in People with Schizophrenia. *Journal of Cognitive Neuroscience*, 1-14.

Lundqvist, M., Herman, P., & Miller, E. K. (2018). Working memory: delay activity, yes! Persistent activity? Maybe not. *Journal of Neuroscience, 38*(32), 7013-7019.

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron, 90*(1), 152-164.

Luo, H., Tian, X., Song, K., Zhou, K., & Poeppel, D. (2013). Neural response phase tracks how listeners learn new acoustic representations. *Current Biology, 23*(11), 968-974.

Luyckx, F., Nili, H., Spitzer, B., & Summerfield, C. (2019). Neural structure mapping in human probabilistic reward learning. *Elife, 8*, e42816.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron, 71*(4), 737-749.

Mahalanobis, P. C. (1936). *On the generalized distance in statistics*.

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron, 56*(3), 530-540.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods, 164*(1), 177-190.

Marshuetz, C. (2005). Order information in working memory: an integrative review of evidence from brain and behavior. *Psychological bulletin, 131*(3), 323.

Masse, N. Y., Hodnefield, J. M., & Freedman, D. J. (2017). Mnemonic encoding and cortical organization in parietal and prefrontal cortices. *Journal of Neuroscience*, 37(25), 6098-6112.

Masse, N. Y., Rosen, M. C., & Freedman, D. J. (2020). Reevaluating the Role of Persistent Neural Activity in Short-Term Memory. *Trends in cognitive sciences, 24*(3), 242-258.

Mendoza-Halliday, D., Torres, S., & Martinez-Trujillo, J. C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature neuroscience*, 17(9), 1255-1262.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of neurophysiology, 100*(3), 1407-1419.

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science, 319*(5869), 1543-1546.

Muller, R. (1996). A quarter of a century of place cells. *Neuron, 17*(5), 813-822.

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences, 114*(2), 394-399.

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing information during working memory: beyond sustained internal attention. *Trends in cognitive sciences, 21*(6), 449-461.

Naya, Y., Chen, H., Yang, C., & Suzuki, W. A. (2017). Contributions of primate prefrontal cortex and medial temporal lobe to temporal-order memory. *Proceedings of the National Academy of Sciences, 114*(51), 13555-13560.

Naya, Y., & Suzuki, W. A. (2011). Integrating what and when across the primate medial temporal lobe. *Science, 333*(6043), 773-776.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

O'keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*: Oxford: Clarendon Press.

Panichello, M. F., & Buschman, T. J. (2020). Selective control of working memory in prefrontal, parietal, and visual cortex. *bioRxiv*.

Parkinson, C., Liu, S., & Wheatley, T. (2014). A common cortical metric for spatial, temporal, and social distance. *Journal of Neuroscience, 34*(5), 1979-1987.

Parthasarathy, A., Herikstad, R., Bong, J. H., Medina, F. S., Libedinsky, C., & Yen, S.-C. (2017). Mixed selectivity morphs population codes in prefrontal cortex. *Nature Neuroscience, 20*(12), 1770-1779.

Parthasarathy, A., Tang, C., Herikstad, R., Cheong, L. F., Yen, S.-C., & Libedinsky, C. (2019). Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nature communications, 10*(1), 1-11.

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science, 321*(5894), 1322-1327.

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. *Proceedings of the National Academy of Sciences, 107*(13), 6046-6051.

Quentin, R., King, J.-R., Sallard, E., Fishman, N., Thompson, R., Buch, E. R., & Cohen, L. G. (2019). Differential brain mechanisms of selection and maintenance of information during working memory. *Journal of Neuroscience, 39*(19), 3728-3740.

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience, 22*(8), 1336-1344.

Rajji, T. K., Zomorrodi, R., Barr, M. S., Blumberger, D. M., Mulsant, B. H., & Daskalakis, Z. J. (2017). Ordering information in working memory and modulation of gamma by theta oscillations in humans. *Cerebral Cortex, 27*(2), 1482-1490.

Roberts, B. M., Hsieh, L.-T., & Ranganath, C. (2013). Oscillatory activity during maintenance of spatial and temporal information in working memory. *Neuropsychologia, 51*(2), 349-357.

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*(6316), 1136-1139.

Schuck, N. W., & Niv, Y. (2019). Sequential replay of nonspatial task states in the human hippocampus. *Science, 364*(6447), eaaw5181.

Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage, 184*, 741-760.

Spaak, E., Watanabe, K., Funahashi, S., & Stokes, M. G. (2017). Stable and dynamic coding for working memory in primate prefrontal cortex. *Journal of Neuroscience, 37*(27), 6503-6516.

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring latent visual working memory representations in human cortex. *Neuron, 91*(3), 694-707.

Stokes, M. G. (2015). 'Activity-silent'working memory in prefrontal cortex: a dynamic coding framework. *Trends in cognitive sciences, 19*(7), 394-405.

Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal

cortex. *Progress in Neurobiology, 184*, 101717.

Sutterer, D. W., Foster, J. J., Adam, K. C., Vogel, E. K., & Awh, E. (2019). Item-specific delay activity demonstrates concurrent storage of multiple active neural representations in working memory. *PLoS biology, 17*(4), e3000239.

Trübutschek, D., Marti, S., Ojeda, A., King, J.-R., Mi, Y., Tsodyks, M., & Dehaene, S. (2017). A theory of working memory without consciousness or sustained activity. *Elife, 6*, e23871.

Tsao, A., Moser, M. B., & Moser, E. I. (2013). Traces of experience in the lateral entorhinal cortex. *Current Biology*, 23(5), 399-405.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M. B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, 561(7721), 57-62.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., . . . Morris, R. G. (2007). Schemas and memory consolidation. *Science, 316*(5821), 76-82.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological bulletin, 138*(6), 1172.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in cognitive sciences, 7*(11), 483-488.

Weisz, N., Kraft, N. G., & Demarchi, G. (2020). Auditory cortical alpha/beta desynchronization prioritizes the representation of memory items during a retention period. *Elife, 9*, e55508.

Whittington, J., Muller, T., Mark, S., Barry, C., & Behrens, T. (2018). *Generalisation of structural knowledge in the hippocampal-entorhinal system.* Paper presented at the Advances in neural information processing systems.

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience, 20*(6), 864.

Wolff, M. J., Kandemir, G., Stokes, M. G., & Akyürek, E. G. (2020). Unimodal and bimodal access to sensory working memories by auditory and visual impulses. *Journal of Neuroscience, 40*(3), 671-681.

Yu, Q., Teng, C., & Postle, B. R. (2020). Different states of priority recruit different neural representations in visual working memory. *PLoS biology, 18*(6), e3000769.

Yue, Q., Martin, R. C., Hamilton, A. C., & Rose, N. S. (2019). Non-perceptual regions in the left inferior parietal lobe support phonological short-term memory: evidence for a buffer account? *Cerebral Cortex, 29*(4), 1398-1413.

Zhou, X., Zhu, D., Qi, X. L., Lees, C. J., Bennett, A. J., Salinas, E., ... & Constantinidis, C. (2013). Working memory performance and neural activity in prefrontal cortex of peripubertal monkeys. *Journal of neurophysiology*, 110(11), 2648-2660.