Behavioral/Cognitive

Decoding Concurrent Representations of Pitch and Location in Auditory Working Memory

Stefan Czoschke,[1,2] Cora Fischer,[1,2] Tara Bahador,[1,2] Christoph Bledowski,[1,2]* and Jochen Kaiser[1,2]*

[1]Institute of Medical Psychology, Medical Faculty, Goethe University, Frankfurt am Main 60528, Germany, and [2]Brain Imaging Center, Medical Faculty, Goethe University, Frankfurt am Main 60528, Germany

Multivariate analyses of hemodynamic signals serve to identify the storage of specific stimulus contents in working memory (WM). Representations of visual stimuli have been demonstrated both in sensory regions and in higher cortical areas. While previous research has typically focused on the WM maintenance of a single content feature, it remains unclear whether two separate features of a single object can be decoded concurrently. Also, much less evidence exists for representations of auditory compared with visual stimulus features. To address these issues, human participants had to memorize both pitch and perceived location of one of two sample sounds. After a delay phase, they were asked to reproduce either pitch or location. At recall, both features showed comparable levels of discriminability. Region of interest (ROI)-based decoding of functional magnetic resonance imaging (fMRI) data during the delay phase revealed feature-selective activity for both pitch and location of a memorized sound in auditory cortex and superior parietal lobule. The latter region showed higher decoding accuracy for location than pitch. In addition, location could be decoded from angular and supramarginal gyrus and both superior and inferior frontal gyrus. The latter region also showed a trend for decoding of pitch. We found no region exclusively coding pitch memory information. In summary, the present study yielded evidence for concurrent representations of pitch and location of a single object both in sensory cortex and in hierarchically higher regions, pointing toward representation formats that enable feature integration within the same anatomic brain regions.

Key words: auditory; decoding; functional magnetic resonance imaging; multivariate pattern analysis; working memory

---

Significance Statement

Decoding of hemodynamic signals serves to identify brain regions involved in the storage of stimulus-specific information in working memory (WM). While to-be-remembered information typically consists of several features, most previous investigations have focused on the maintenance of one memorized feature belonging to one visual object. The present study assessed the concurrent storage of two features of the same object in auditory WM. We found that both pitch and location of memorized sounds were decodable both in early sensory areas, in higher-level superior parietal cortex and, to a lesser extent, in inferior frontal cortex. While auditory cortex is known to process different features in parallel, their concurrent representation in parietal regions may support the integration of object features in WM.

Introduction

Working memory (WM) is a capacity-limited system enabling the temporary maintenance and manipulation of information (Baddeley, 1992). Functional magnetic resonance imaging (fMRI) has been used to identify brain regions underlying the storage of WM contents. In line with monkey electrophysiology and lesion studies (Fuster and Alexander, 1971; Petrides, 2000), univariate analyses have supported the role of prefrontal and parietal cortex for stimulus processing in WM (Courtney et al., 1997; Pessoa et al., 2002; Curtis and D'Esposito, 2003; Todd and Marois, 2004; Bledowski et al., 2009). Multivariate decoding approaches, allowing the identification of content-specific activation patterns (Kriegeskorte et al., 2006; Haynes, 2015), have shown that early sensory regions are involved in the temporary storage of low-level sensory features (Harrison and Tong, 2009; Serences et al., 2009; Christophel et al., 2012; Riggall and Postle, 2012). Stimulus-specific information was decodable also from the intraparietal sulcus (Bettencourt and Xu, 2016; Galeano Weber et al., 2016; Christophel et al., 2018; Lorenc et al., 2018; Rademaker et al., 2019) and even frontoparietal regions (Ester et al., 2015). In summary, representations of WM contents are distributed across multiple cortical areas, whereby hierarchically
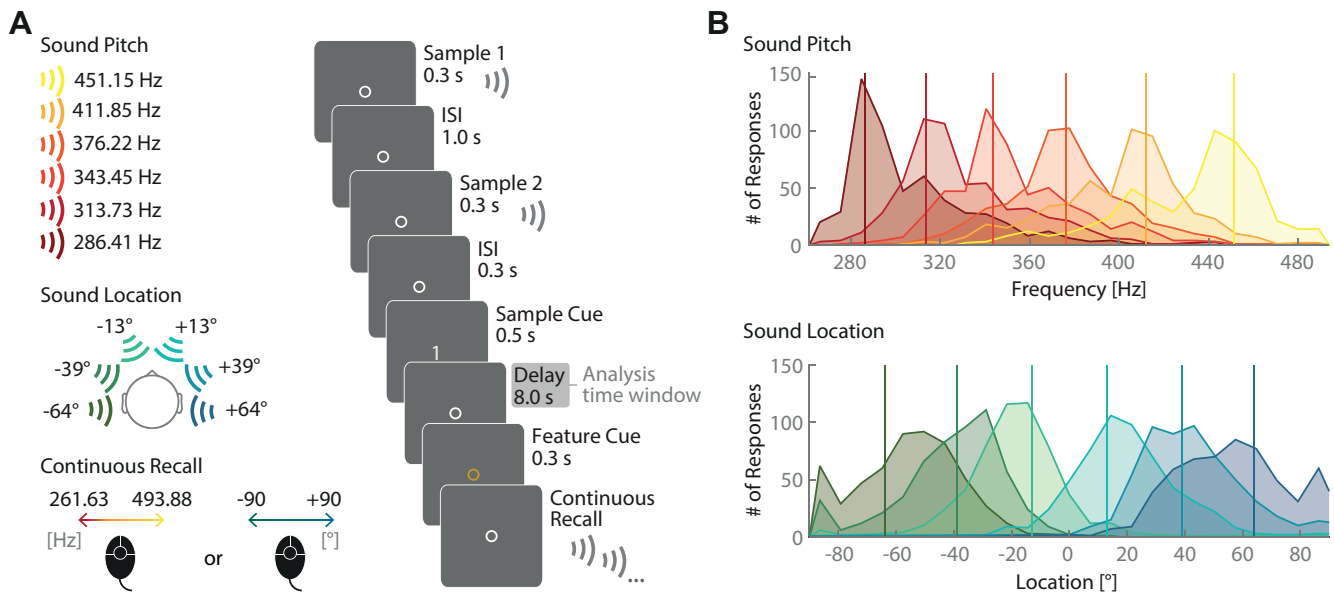
Figure 1. Auditory WM task and behavioral data. **A**, left panel, Pitch and location feature values of complex sound samples presented in the WM task. Bottom left panel, Range of the feature values used during the continuous recall of pitch and location. Right panel shows the trial structure: participants heard two sequentially presented sound samples and were visually cued to remember the precise pitch and location of a single target sound. After an 8-s delay, participants were cued to recall either pitch or location of the cued target sound by adjusting the corresponding feature of a "pulsating" probe sound using a trackball. **B**, Response distributions for each recalled pitch (upper panel) and location (lower panel) of the target sound sample.

higher regions likely code information at a higher level of abstraction than early sensory areas (Christophel et al., 2017).

Although multiple items can be held in visual WM (Ma et al., 2014), most multivariate decoding studies have required the maintenance of one single-feature item only (Harrison and Tong, 2009; Rademaker et al., 2019), or they have assessed the classification performance for one item when additional stimuli had to be memorized (Emrich et al., 2013; Gosseries et al., 2018). While some investigations have suggested that only a single item in the focus of attention is accompanied by a stimulus-selective activation pattern (Lewis-Peacock et al., 2012; LaRocque et al., 2017; Wolff et al., 2017), others have demonstrated the decodability of currently irrelevant (Peters et al., 2015; van Loon et al., 2018) or unattended contents (Christophel et al., 2018). When up to two visual objects had to be maintained, concurrent decoding of their positions revealed an inverse relationship between memory load and classification performance (Sprague et al., 2014). While these previous studies have demonstrated concurrent decoding of the same target feature in two memorized objects, it remains unclear whether and where in the brain two different, task-relevant features of one perceptual object can be decoded concurrently during maintenance in WM.

While most WM decoding studies have investigated visual materials, less is known about the maintenance of auditory stimulus features. The few existing multivariate studies have focused on sound identity defined either by frequency or amplitude modulation rate, which could be decoded in auditory regions (Linke et al., 2011) and, additionally, in inferior frontal (Kumar et al., 2016) or precentral cortex (Uluç et al., 2018). In contrast, we are not aware of multivariate analyses of WM maintenance of sound location, which is thought to be processed in parallel to sound identity along a separate cortical pathway (Rauschecker, 1998; Alain et al., 2001; Arnott et al., 2004).

The aims of the present study were thus twofold. We sought to assess the concurrent decoding of two features of a single memorized object and to expand the knowledge on WM representations of acoustic stimulus features. We hypothesized that both pitch and location of

a memorized sound can be decoded concurrently from auditory cortex, whereas regions along the putative auditory dorsal and ventral streams should preferentially code either location or pitch, respectively.

Materials and Methods

Participants
A total of 42 healthy participants (29 females, mean age 22.7 years, SD = 3.96) with normal or corrected-to-normal vision were recruited for a behavioral pilot test that comprised the experimental fMRI tasks and served as a performance screening and practice session. Fourteen participants were excluded after behavioral testing for one of the following reasons: reporting the task as being very difficult (two participants), detection threshold for location changes exceeding 70° (six participants), poor accuracy in reproducing pitch or location from memory (three participants), drop-out before the first fMRI session (three participants); 28 participants (20 females, mean age 21.04 years, SD = 5.39) completed three fMRI sessions lasting ~2 h each. All participants provided written informed consent and received monetary reimbursement for participation (€10/h). The study was conducted at the Goethe University and was approved by the ethics committee of the University of Frankfurt medical faculty.

Stimuli and apparatus
The stimuli were two-dimensional feature combinations of a complex sound and a spatial location (Fig. 1A). The spectral frequencies of the presented sounds were a combination of three harmonic band-passed noises (a fundamental frequency and two harmonics, each bandpass filtered to a bandwidth of 1/10 octave). Sound duration was 300 ms with rise and fall times of ~50 ms. The fundamental frequencies ranged from 286.41 to 451.15 Hz in six equal steps in logarithmic space. The spatial stimulus feature was generated by introducing an interaural time difference (i.e., a sound onset delay between the left and right headphone) of 0.53, 0.34, or 0.12 ms, corresponding roughly to 64°, 39°, and 13° from center to the left and to the right, respectively. The stimulus pool thus consisted of 36 (6 × 6) different combinations of pitch and location. Stimuli were processed with an external soundcard (Fireface UC, 192 kHz sampling rate, RME) and presented at a comfortable intensity (~85–95 dB SPL) via MRI-compatible noise cancellation headphones (OptoActive, Optoacoustics Ltd). Stimulus construction and timing were controlled with MATLAB R2012b (The MathWorks) and the Psychophysics Toolbox (Brainard, 1997).

Experimental design: localizer experiment

In the localizer experiment, participants had to selectively encode the pitch or location of two sequentially presented sounds (300-ms sound duration each) and compare one of them, which was retrospectively cued, to a subsequent test sound. Participants had to decide whether the test sound matched the cued item on the task-relevant feature dimension (two-alternative forced choice match/non-match task) within a response window of 2 s. No specific instructions were given with respect to speed or accuracy. The task-relevant feature was altered between recording blocks and was indicated by a colored (yellow or blue) fixation circle, which was presented at the screen center for 1 s at block onset, and remained on the screen throughout the block (color × task association was counterbalanced across participants). Participants were instructed to maintain central fixation throughout the experimental blocks. Each block lasted 16 s and comprised three trials followed by a 16-s resting period. Each trial began with the presentation of the first stimulus for 300 ms, followed by an interstimulus interval of 300 ms and the presentation of the second stimulus for 300 ms; 300 ms after the second stimulus, a numeric cue appeared for 500 ms that indicated whether the first or second stimulus had to be compared with the upcoming test tone. After cue presentation, the test stimulus appeared immediately for 300 ms followed by a 1.7-s response period in which the participants indicated via button press whether the test stimulus was identical to the target stimulus or deviated from it on the task-relevant feature dimension. If participants did not respond within 2 s, the trial was recorded as incorrect. Then a visual feedback (red or green fixation circle for incorrect or correct responses, respectively) appeared for 300 ms, followed by an intertrial interval of 1 s. There was one run per fMRI session. Each run contained 48 blocks (24 per feature), resulting in 72 trials per run and feature. Half of the trials per feature were match trials (i.e., target and test stimulus were identical). In non-match trials the difference between the target and the test stimulus was controlled by a one-up/two-down staircase procedure targeting 70.7% correct non-match responses. The initial difference between target and test was 80° for location trials and 0.04 $\log_{10}$(Hz) for pitch trials [with a step size of 5° or 0.02 $\log_{10}$(Hz) for location and pitch, respectively]. The staircases ran continuously across all four sessions (behavioral session and three fMRI sessions). On each trial, pitch and location were drawn randomly without replacement for both stimuli, with the restriction that laterality was counterbalanced per run and target feature. Serial position and laterality of the target stimulus were counterbalanced per run and target feature. Please note that data from the localizer experiment were also used for a separate study (Erhart et al., 2021).

Experimental design: main experiment

In the main experiment, participants memorized pitch and location of two sequentially presented sounds for a continuous recall task that allowed to freely adjust a probe stimulus to reproduce the pitch or location value from memory (Fig. 1A). After stimulus presentation, participants were informed by a sample cue which stimulus (but not which feature) would be task-relevant at the end of the trial. Thus, participants had to maintain both pitch and location of the cued stimulus throughout the delay period. After the retention period, participants were informed via a feature cue about the to-be-reported feature (pitch or location) and subsequently reported the target feature via a continuous probe-adjustment procedure. The trial structure was as follows: 1-s fixation period with a white fixation circle at the screen center, first sample stimulus (300 ms), interstimulus interval (1 s), second sample stimulus (300 ms). After a 300-ms delay, a sample cue (the number 1 or 2) appeared for 500 ms that indicated the serial position of the task-relevant item, i.e., whether the first or the second sample stimulus had to be maintained. The 8-s delay period was followed by a feature cue (500 ms) that indicated by color (yellow or blue) which feature to report in the subsequent continuous recall procedure. After the feature cue, a probe tone was presented. The probe tone lasted 300 ms and was continuously repeated either until a response was made or the 10-s response window ended. The probe tone had a random start value on the to-be-reported feature dimension and was identical to the target stimulus on the task-irrelevant feature dimension. Participants adjusted the respective target feature by rotating a trackball to the left or right and entered their response via button press. The response space for pitch was a continuous space that

spanned one octave (C4 to B4) from 261.62 to 493.88 Hz. The response space for location was a continuous space ranging from 90° on the left to 90° on the right with a maximum interaural time difference of 0.68 ms. Participants received feedback about their recall accuracy at the end of each trial. Following the response, the fixation circle took on a color between green and red indicating the error magnitude (300 ms). The intertrial interval (blank screen) lasted 3, 4, 5, or 6 s. The intertrial interval duration was counterbalanced and randomized across trials.

As we had expected a relatively poor classifier performance on the basis of comparable previous studies (Kumar et al., 2016), we decided to record a large number of trials per condition requiring three separate MR recording sessions per participant. Each fMRI session contained four runs of the experiment. Each run contained 24 trials (96 trials per session). Within each session, each of the 36 feature combinations appeared at least once as the target and once as the non-target for each serial position, resulting in 72 unique examples of feature combination × serial position for target and non-target stimuli. The remaining 24 trials per session were filled up with randomly drawn feature combinations. Serial position and retrieval feature of the target stimulus were counterbalanced within runs.

Data acquisition

fMRI data were collected with a 3-Tesla Magnetom Prisma MR scanner (Siemens), located at the Brain Imaging Center Frankfurt. We used a 64-channel head coil. We acquired whole-brain echo-planar images (EPI; 51 axial slices aligned approximately in parallel to the anterior and posterior commissures, 3 × 3 × 2 mm resolution, 1-mm gap, field of view (FoV) = 192 mm, repetition time (TR) = 1 s, time echo (TE) = 30 ms, flip angle = 90°, interleaved acquisition) and a high-resolution T1-weighted image (192 sagittal slices, 1 × 1 × 1 mm resolution, FoV =256 mm, TR = 1 s, TE = 2.52 ms). Within each session, we collected four runs of the WM main experiment, followed by the T1 image acquisition, and the localizer task. Between two consecutive runs, and before the localizer task, we recorded five EPI volumes with reversed phase encoding direction for the correction of field inhomogeneities.

fMRI preprocessing

Functional imaging data were preprocessed using FSL (https://fsl.fmrib.ox.ac.uk) and SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12). Functional data were motion-corrected via FSL MCFLIRT. A 52nd slice was added to each functional volume by duplicating the bottom slice of each volume to perform distortion correction (this was necessary, as FSL's topup procedure for distortion correction requires an even slice number). Subsequently, all runs were distortion corrected via FSL's topup (for each functional run a susceptibility-induced off-resonance field was estimated based on two consecutively acquired images with reversed phase-encoding blips, and applied to the functional images for distortion correction (Andersson et al., 2003; Smith et al., 2004). After distortion correction, the 52nd slice was removed and all distortion-corrected runs were then realigned onto the first volume of the first session's localizer run (as the localizer was closest to the T1-image acquisition) and, in a second step, onto the mean image of all runs, using SPM. No normalization, spatial smoothing or slice-time correction were performed. The anatomic image was co-registered to the functional mean image.

Univariate analyses: localizer task

To identify feature-selective voxels for the multivariate decoding analysis, we estimated the BOLD activity during pitch and location blocks of the localizer task. A general linear model (GLM) with eight regressors (pitch blocks, location blocks and six motion regressors) was designed using hemodynamic response functions (HRFs) time-locked to the onset of each block with a duration of 16 s. The GLM included the localizer runs of all three fMRI sessions. To identify voxels that responded to pitch or location processing, we generated separate t-maps for pitch and location blocks by contrasting each regressor with the implicit baseline. These t-maps were subsequently used for specification of the regions of interest (ROIs; see next section).

ROIs

ROIs were based on anatomically defined regions that had been found to be involved in the processing of auditory spatial and non-spatial information (Arnott et al., 2005; Alain et al., 2008, 2018). Specifically, we predefined

Table 1. ROI Sizes and proportion of voxels for feature-selective and overlapping areas across ROIs

ROI	ROI size				Selectivity		
	Location		Pitch		Location	Pitch	Overlap
	Voxels	N	Voxels	N	%	%	%
Auditory cortex	700	27	200	28	71	< 1	29
	600	1					
Temporal pole	600	23	100	27	84	< 1	16
	1200	5	400	1			
Inferior frontal	600	28	500	28	23	8	68
Inferior parietal	1700	26	200	26	87	< 1	13
	1600	2	400	1			
			600	1			
Superior frontal	800	21	1200	18	36	17	47
	1600	7	200	10			
Superior parietal	1100	28	1000	16	22	5	73
			800	12			
Middle frontal	500	13	1100	12	36	4	60
	400	10	1300	7			
	1000	3	1900	5			
	300	2	1700	2			
			100	1			
			1600	1			

ROI size indicates the number of voxels included in the decoding analysis for each feature. Different numbers of voxels were selected for subgroups of participants. Selectivity refers to the group mean percentage of feature-selective and overlapping ROI areas included in the decoding analysis. To quantify the percentage of unique and overlapping voxels used for location and pitch decoding, we combined the final ROI maps for pitch and location within each subject and calculated the proportion of voxels exclusively used for location decoding, exclusively used for pitch decoding and used for both. N, number of participants.

a set of seven ROIs using the *Harvard-Oxford Cortical and Subcortical Structural Atlases* (Desikan et al., 2006). The auditory cortex ROI included Heschl's gyrus, planum temporale and the posterior division of the superior temporal gyrus, the inferior parietal ROI comprised the angular gyrus and the anterior and posterior division of the supramarginal gyrus, the superior frontal ROI comprised the superior frontal gyrus, the temporal pole ROI comprised the anterior divisions of the superior temporal and middle temporal gyrus and the temporal pole, the inferior frontal ROI comprised the pars triangularis and pars opercularis of the inferior frontal gyrus, the superior parietal ROI comprised the superior parietal lobule, and the middle frontal ROI comprised the middle frontal gyrus. We considered the auditory cortex ROI as the starting point of both auditory processing streams, the inferior parietal and superior frontal ROIs were thought to be associated with the auditory dorsal pathway, and the temporal pole and inferior frontal ROIs to be associated with the auditory ventral pathway. The superior parietal and middle frontal ROIs were chosen as higher-level WM-relevant regions not predominantly associated with one of the pathways. The anatomic probability maps (MNI space) were extracted from the FMRIB Software Library (FSL) *Harvard-Oxford Cortical and Subcortical Structural Atlases* (Desikan et al., 2006). ROIs were transformed into each participant's native space using SPM's segmentation and normalization functions. Finally, maps were thresholded to exclude voxels with a probability of forming part of the chosen areas of <0.1, which at the same time avoided overfitting (Christophel et al., 2018). To further condense the ROIs to feature-selective voxels, each of the binary ROI maps was multiplied with the pitch and location contrast t-maps from the localizer task. This resulted in separate versions of ROI maps for pitch and location for each participant. We then created 30 different versions of each ROI and feature by selecting the n voxels with the strongest response above or below baseline, i.e., with the n largest absolute t values. N varied from 100 to 3000 voxels in steps of 100 voxels (Christophel et al., 2018). Multivariate analysis was performed on each of these ROIs and then subjected to a cross-validation scheme to select the final ROI size for each subject independently. For details, see below, Multivariate analyses: main experiment.

Multivariate analyses: main experiment
Multivariate analyses were conducted on session-specific delay-period t-maps for each of the 36 feature combinations × serial position of the target and non-target stimulus, respectively. First, trial-specific BOLD signals for the delay period were estimated using the least squares single

(LS-S) approach described by Mumford et al. (2012). To estimate trial-specific delay period activity, we generated one GLM per trial. Each GLM contained five condition regressors (one regressor for the single trial's delay period, one regressor for the delay period of the remaining trials of the run, and three regressors for the encoding period, stimulus cue and response period across all trials of the run) and six motion regressors. Each condition regressor was convolved with the HRF and time-locked to the event onset. t-Maps for the delay period of each trial were then calculated by contrasting the respective trial-specific delay regressor with the implicit baseline. The resulting 96 trial-wise t-maps per recording session were then averaged into 72 condition examples (36 feature combinations × 2 serial positions) for target and non-target stimuli, respectively. Multivoxel pattern analysis (MVPA) classification was done with the MVPA-light toolbox for MATLAB (Treder, 2020). We used linear SVM classification to decode pitch and location information within each ROI separately. Data were median-split into two classes for each feature dimension. Our decoding analysis pursued two goals. First, to decode stimulus-specific information of the task-relevant item. Second, to demonstrate that the decoded signal was specific to the prioritized, task-relevant item and did not reflect memory-unspecific, encoding related signals. To this end, our classification models were exclusively trained on data with target-stimulus class labels but separately applied to classify target and non-target feature classes. The degree to which decoding accuracy differs between target and non-target classification reflects signal differences that are introduced only by the retrocue that indicated the to-be-recalled item and thus assigned the roles of target and non-target. The classification models were iteratively trained on the single-subject data of two sessions and tested on data of the left-out third session (leave-one-session-out) to guarantee independence of training and test data. Classification accuracy was then determined as the mean accuracy of the three train-test cycles. This procedure was performed separately for classification of pitch and location within each ROI and voxel count. To estimate and select the optimal number of voxels for each ROI and feature per participant, we used a leave-one-participant-out nested cross-validation procedure that determined the voxel counts per ROI for each participant based on the group statistics after exclusion of the participant in accordance with Christophel et al. (2018). This procedure optimized decoding accuracies for each region and feature independently, thereby increasing comparability across features. In contrast to procedures with fixed, predetermined voxel counts or fixed activity thresholds, it required only minimal prior assumptions about location-
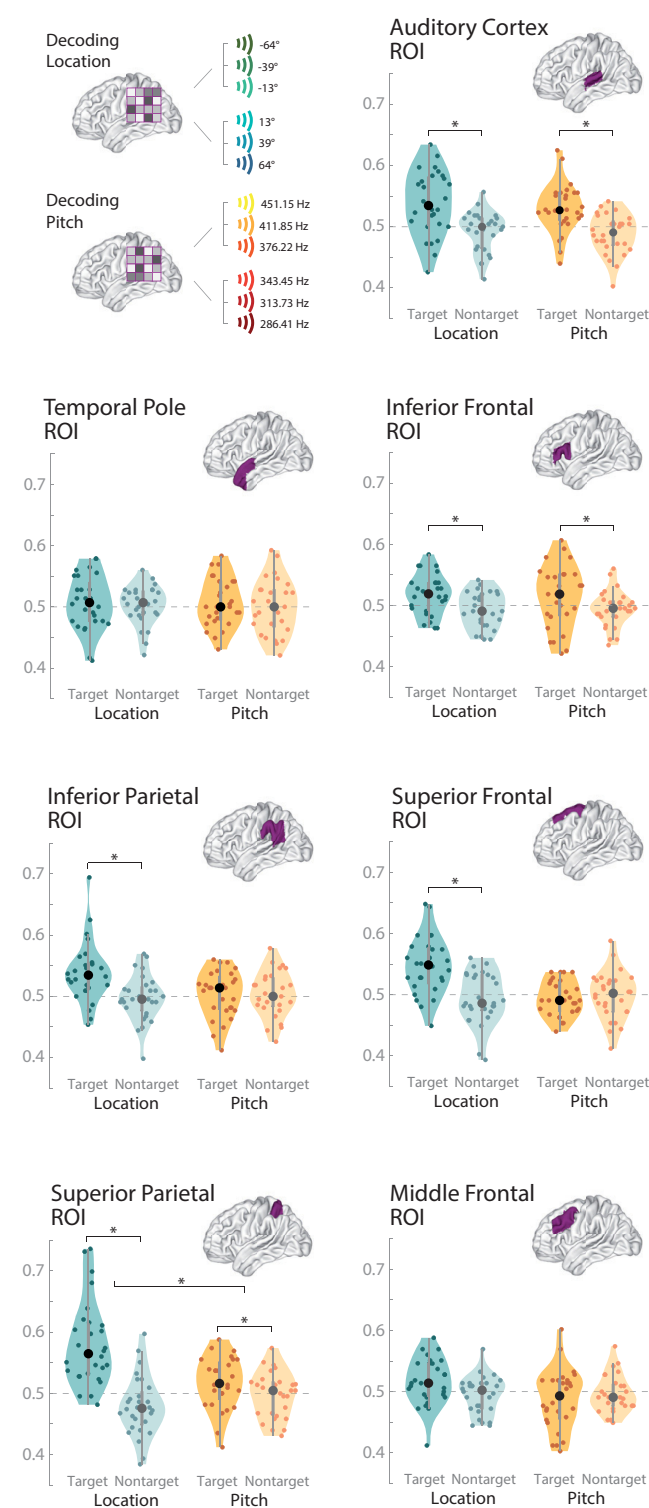
Figure 2. Concurrent decoding accuracy of location and pitch of a sound sample retained in auditory WM. Upper left panel, Simplified illustration of the decoding procedure (for details, see Materials and Methods). Remaining panels, Decoding accuracy of pitch and location of the target (i.e., the to-be-remembered sample) and non-target sound samples during the delay phase of the WM task, in seven ROIs. Colored dots represent the values of individual participants, the black and dark gray dots reflect the group mean, the bold gray bars indicate the interquartile range, the gray thin bars represent the whiskers with maximum 1.5 interquartile range and shading represents the density trace. Each panel contains a rendered representation of the human brain depicting the corresponding ROI. Asterisks indicate significant differences between conditions; for details, see "Results. We found main effects of task-relevance (target vs non-target) in the auditory cortex and inferior frontal ROIs. The inferior parietal and superior frontal ROIs showed main effects of feature accompanied by an

specific or pitch-specific signaling, and it avoided double dipping in voxel selection. Specifically, for each participant and ROI, we excluded the participant's data, averaged the classification accuracy of the remaining participants for each of the 30 voxel counts of the ROI, and determined the ROI size with the highest mean classification accuracy. This ROI size was then used to select the classification accuracy for the ROI of the left-out participant. This procedure was performed for pitch and location classification separately. ROI sizes were determined on the classifier performance for target feature classification. Table 1 lists the resulting ROI sizes and the spatial overlap for feature-selective voxels across anatomic ROIs.

Statistical analysis
Behavioral data
To estimate the accuracy of recall from WM, we calculated the behavioral distinctiveness of the memory representations by computing the mean and standard deviation of the behavioral response distributions for each of the six presented feature values per feature dimension. Subsequently, Cohen's d (mean difference divided by the pooled standard deviation) was calculated for neighboring feature values (e.g., 64° left vs 39° left; 39° left vs 13° left; and so on) and averaged across the feature dimension. This resulted in a mean behavioral distinctiveness measure for pitch and location, respectively.

fMRI decoding
All analyses were conducted at the ROI level. An α level of 0.05 was set for all analyses with no correction of multiple comparisons across ROIs. Within each ROI we compared the decoding accuracies with a repeated-measures ANOVA that comprised the factors task-relevance (target vs non-target) and feature (location vs pitch). The task-relevance comparison reflected whether the decoded signal carried memory-specific information unique to the prioritized target item. The feature comparison reflected the relative decoding sensitivity for location and pitch information in the respective ROI. Additionally, we calculated *post hoc* sign permutation tests for every comparison that yielded an effect of task-relevance to test for above-chance decoding. For sign permutation tests the null distribution of group means (10,000 iterations) was generated by subtracting the chance level (0.5) from individual decoding accuracies, randomly inverting the sign of the resulting differences and computing the group mean. We then calculated the proportion of simulated group means equal or larger than the empirical group mean of the decoding accuracies and reported this proportion as the p value.

Results
Behavioral performance
The response distributions for each of the six presented feature values for pitch and location are depicted in Figure 1B. Albeit showing some overlap, the reproduced feature values were clearly separable based on their respective behavioral response profiles. While pitch responses appeared continuously distributed across the feature dimension, location showed a lateralized pattern with fewer responses around the center of the location space. Behavioral distinctiveness for location was $d = 1.16$ (SD = 0.29) and for pitch it was $d = 0.98$ (SD = 0.73). By analogy to Cohen's d, these values can be interpreted as reflecting a strong average effect size (i.e., a low degree of overlap between neighboring response distributions). The behavioral distinctiveness did not differ between both features (paired t test, $t_{(27)} = 1.31$, $p = 0.200$), indicating that the discriminability of the presented pitches and locations was comparable.

To assess whether behavioral performance remained stable across sessions, we calculated additionally a repeated-measures ANOVA with the factors feature × session to analyze changes in

←

interaction between task-relevance and feature. Main effects of task-relevance and feature and an interaction between both factors were found in the superior parietal ROI.

stimulus discriminability. The ANOVA showed no significant effects (main effect of feature, $F_{(1,27)} = 1.01$, $p = 0.323$; main effect of session, $F_{(2,54)} = 0.64$, $p = 0.531$; interaction, $F_{(2,54)} = 0.08$, $p = 0.925$), indicating a constant level of memory performance. The behavioral distinctiveness scores for Location were $d = 1.23$ (SD = 0.28, range = 0.73–1.90) for session 1, $d = 1.33$ (SD = 0.37, range = 0.58–2.17) for session 2 and $d = 1.29$ (SD = 0.41, range = 0.37–2.16) for session 3. For pitch, behavioral distinctiveness amounted to $d = 1.10$ (SD = 0.85, range = 0.49–4.54) for session 1, $d = 1.14$ (SD = 1.27, range = 0.20–6.95) for session 2 and $d = 1.14$ (SD = 0.65, range = 0.33–2.83) for session 3.

fMRI decoding

We compared the effects of task-relevance and feature on the decoding accuracy using repeated-measures ANOVAs for each ROI separately. The results are shown in Figure 2.

Auditory cortex. In the auditory cortex ROI, decoding of the task-relevant features was more accurate than of the task-irrelevant features, indicating memory-specific signal for pitch and location. We observed a main effect of task-relevance ($F_{(1,27)} = 33.75$, $p < 0.001$), no main effect of feature ($F_{(1,27)} = 0.85$, $p = 0.366$), and no interaction ($F_{(1,27)} = 0.07$, $p = 0.795$). Decoding of both task-relevant location and task-relevant pitch were significantly above chance (both $p < 0.001$, post hoc sign permutation tests).

Auditory dorsal stream. In the inferior parietal ROI, decoding of the task-relevant location was more accurate than decoding of the task-irrelevant location, indicating memory-specific signal for location information. Decoding of task-relevant location was also more accurate than of task-relevant pitch, indicating greater sensitivity to information of the target's location relative to pitch information. We observed a main effect of task-relevance ($F_{(1,27)} = 5.07$, $p = 0.033$), a trend for a main effect of feature ($F_{(1,27)} = 4.05$, $p = 0.054$), and a significant interaction ($F_{(1,27)} = 7.14$, $p = 0.013$). Post hoc paired t tests revealed that the effect of task-relevance was specific to location decoding ($t_{(27)} = 3.23$, $p = 0.002$). Decoding of the target location [mean (M) = 53.85%, SD = 4.94] was more accurate than non-target location decoding (M = 49.75%, SD = 3.73). Decoding of the target location was significantly above chance ($p < 0.001$, post hoc sign permutation test). In contrast, there was no effect of task-relevance for pitch (target: M = 50.35%, SD = 3.80, non-target: M = 50.33% SD = 3.64, $t_{(27)} = 0.02$, $p = 0.493$), indicating a lack of decodable memory-specific signal for pitch information. Furthermore, we found a significant effect of feature for the target features. Decoding of the target location was significantly more accurate than decoding the target pitch ($t_{(27)} = 3.17$, $p < 0.004$). Decoding of the non-target features did not differ significantly ($t_{(27)} = -0.59$, $p = 0.559$).

In the superior frontal ROI, decoding of the task-relevant location was more accurate than decoding of the task-irrelevant location, indicating memory-specific signal. Decoding of the task-relevant location was also more accurate than of the task-relevant pitch, indicating greater sensitivity to information about the target's location relative to pitch information. Decoding of the task-relevant pitch did not differ from that of the task-irrelevant pitch, indicating that there was no decodable memory-specific signal for pitch information. We observed a main effect of task-relevance ($F_{(1,27)} = 9.14$, $p = 0.005$), a main effect of feature ($F_{(1,27)} = 10.81$, $p = 0.003$), and an interaction ($F_{(1,27)} = 14.09$, $p < 0.001$). Post hoc paired t tests revealed that the effect of task-relevance was specific to location decoding ($t_{(27)} = 4.29$, $p < 0.001$). Decoding of the target location (M = 54.72%, SD = 4.64) was more accurate than non-target location decoding (M =

49.41%, SD = 4.17). Decoding of the target location was significantly above chance ($p < 0.001$, post hoc sign permutation test). There was no effect of task-relevance for pitch (target: M = 49.43%, SD = 2.86, non-target: M = 49.85% SD = 3.86, $t_{(27)} = -0.44$, $p = 0.667$). Furthermore, the effect of feature was specific to target features. Decoding of the target location was significantly more accurate than decoding of the target pitch ($t_{(27)} = 4.78$, $p < 0.001$). Decoding of the non-target features did not differ significantly ($t_{(27)} = -0.43$, $p = 0.671$).

Auditory ventral stream. In the temporal pole ROI, we observed no main effects of task-relevance ($F_{(1,27)} = 0.35$, $p = 0.562$) or feature ($F_{(1,27)} = 0.01$, $p = 0.905$), and no interaction ($F_{(1,27)} = 0.03$, $p = 0.859$), indicating a lack of decodable memory-specific signals for pitch or location. In the inferior frontal ROI, decoding of the task-relevant features was more accurate than decoding of the task-irrelevant features, indicating memory-specific signal for pitch and location. We observed a main effect of task-relevance ($F_{(1,27)} = 7.58$, $p = 0.010$), but no main effect of feature ($F_{(1,27)} = 0.02$, $p = 0.888$), and no interaction ($F_{(1,27)} = 0.10$, $p = 0.756$), i.e., there was no significant difference between pitch and location. Post hoc analysis showed that decoding of task-relevant location was significantly above chance ($p = 0.005$, post hoc sign permutation test), while there was a trend for above-chance decoding of task-relevant pitch ($p = 0.073$, post hoc sign permutation test).

Higher-level WM regions. In the superior parietal ROI, decoding of the task-relevant features was more accurate than decoding of the task-irrelevant features, indicating memory-specific signal for pitch and location. Moreover, decoding of the task-relevant location was more accurate than decoding of the task-relevant pitch, indicating greater sensitivity to information of the target's location relative to pitch information. We observed a main effect of task-relevance ($F_{(1,27)} = 30.21$, $p < 0.001$), a main effect of feature ($F_{(1,27)} = 6.23$, $p = 0.019$), and an interaction ($F_{(1,27)} = 30.93$, $p < 0.001$). Post hoc paired t tests showed that the effect of task-relevance was significant for both location decoding ($t_{(27)} = 6.53$, $p < 0.001$) and pitch decoding ($t_{(27)} = 1.88$, $p = 0.035$). Decoding of the target location (M = 58.01%, SD = 6.83) was more accurate than non-target location decoding (M = 47.91%, SD = 4.84). Decoding of the target pitch (M = 51.69%, SD = 4.19) was more accurate than non-target pitch decoding (M = 49.74%, SD = 3.73). Decoding of both task-relevant location and task-relevant pitch were significantly above chance ($p < 0.001$, $p = 0.022$, respectively, post hoc sign permutation tests). Furthermore, the effect of feature was more pronounced for target than non-target features. Decoding of the target location was significantly more accurate than decoding of the target pitch ($t_{(27)} = 4.65$, $p < 0.001$). There was a trend for decoding of the non-target features ($t_{(27)} = -2.00$, $p = 0.056$). In the middle frontal ROI, decoding of location information was more accurate than decoding of pitch information, indicating a greater sensitivity to location information. This effect, however, was not memory-specific. We observed no main effect of task-relevance ($F_{(1,27)} = 0.63$, $p = 0.434$), but a significant main effect of feature ($F_{(1,27)} = 7.33$, $p = 0.012$), and a trend for an interaction ($F_{(1,27)} = 3.97$, $p = 0.056$).

In summary, our repeated-measures ANOVAs revealed decoding of to-be-remembered memory content for both pitch and location in auditory cortex, inferior frontal cortex and superior parietal lobule, with a higher decoding accuracy for location than pitch in the latter region. In addition, location could be decoded from inferior parietal and superior frontal cortex. We found no region exclusively coding pitch memory information.

Discussion

The present study investigated the concurrent maintenance of sound location and pitch in auditory WM. Using an fMRI decoding paradigm, we found stimulus-specific activation patterns during the delay phase of a WM task in cortical regions along the putative auditory dorsal and ventral processing streams and in higher-level frontoparietal areas. As expected, both location and pitch could be decoded from auditory cortex. We had also expected that pitch was decodable in auditory ventral stream regions like inferior frontal cortex, where we found higher decoding accuracy for the cued target and a trend for decoding above chance. As hypothesized, sound location was decoded from regions of the auditory dorsal stream including inferior parietal and superior frontal cortex. Unexpectedly, location could also be decoded in the inferior frontal ROI. Both location and, to a lesser extent, pitch were decodable from the superior parietal ROI. We failed to find pitch-specific activation patterns in the temporal pole ROI.

The present study showed that two task-relevant features of one acoustic object could be decoded concurrently during maintenance in the delay phase of a WM task. So far, concurrent decoding of different stimulus attributes has been shown in perceptual paradigms only. In the visual modality, concurrent representations of object identity and location have been found in lateral occipital cortex (Cichy et al., 2011), and facial gender and affect were decodable from sensory cortex and attention networks (Long and Kuhl, 2018). In the auditory domain, independent representations have been reported for sound frequency and amplitude modulation in the superior temporal plane (Sohoglu et al., 2020). Considering WM, concurrent decoding of multiple visual contents has been restricted to the same feature of different perceptual objects (Sprague et al., 2014). More recent studies using oriented gratings (Christophel et al., 2018) or object categories (van Loon et al., 2018) have shown decoding of currently unattended items in parallel with attended ones. Similarly, in a previous MVPA study we found that attending one position enhanced the discriminability of an unattended position on the same object (Peters et al., 2015). In the present study participants had to focus their attention on two features of the same memorized object, which were decodable both in sensory and higher-order brain regions. The demonstration of concurrent decodability of different object features opens the possibility to test, e.g., whether attending to one task-relevant feature of an object elicits spontaneous representations of other, task-irrelevant features of the same object. For example, task-irrelevant spatial positions of different types of visual memoranda could be decoded from electroencephalographic alpha band activity (Foster et al., 2017). While this result reflected the importance of location for visual object formation (van Ede et al., 2019; Fischer et al., 2020), given the different nature of visual and auditory objects, it is conceivable that memorized auditory object features would be accompanied by spontaneous representations of sound identity-related attributes like pitch (Kubovy and Van Valkenburg, 2001).

Our findings provide only partial support for the WM maintenance of spatial and non-spatial sound features in regions of the putative auditory dorsal and ventral streams, respectively (Rauschecker, 1998; Arnott et al., 2004). A WM representation of pitch in the auditory cortex ROI replicated previous fMRI decoding studies (Linke et al., 2011; Kumar et al., 2016). This finding is also consistent with the sensory recruitment hypothesis (D'Esposito and Postle, 2015). We also found evidence for pitch representation in inferior frontal cortex, as reflected by the task-relevance main effect without a Task relevance × feature

interaction and a trend for above-chance decoding in the *post hoc* permutation test, which is consistent with both multivariate (Kumar et al., 2016) and some univariate fMRI studies (Alain et al., 2001; Gaab et al., 2003). The failure to find pitch decoding in the temporal pole ROI could be attributable to task-dependent modulations of higher-level auditory regions. Previous fMRI studies have found that the anterior superior temporal gyrus was selectively involved in perceptual discrimination but not WM tasks for both pitch and location, whereas inferior parietal cortex showed the opposite response pattern (Rinne et al., 2009, 2012). While we are not aware of any decoding studies of memorized spatial sound attributes, feature-selective activity patterns for sound location in the inferior parietal and superior frontal ROIs were in line with previous univariate fMRI research (Alain et al., 2001, 2008; Rämä et al., 2004; Leung and Alain, 2011).

In addition, we had selected a middle frontal and a superior parietal ROI as higher-level regions with a known relevance for WM. Here, we found that the superior parietal ROI contained feature-specific information about both location and pitch. Visual WM research has demonstrated the contribution of the superior parietal lobule to WM capacity limitations (Todd and Marois, 2004; Xu and Chun, 2006) and its involvement in attentional prioritization (Nobre et al., 2004; Bledowski et al., 2009). The involvement of frontoparietal attention systems is consistent with previous electroencephalographic studies showing modulations of event-related potentials and alpha power when target stimuli are attentionally selected in AWM tasks (Backer et al., 2015; Lim et al., 2015). Feature-specific neural signatures as reported for sound locations versus categories (Backer et al., 2015) could underlie the present decodability of spatial and non-spatial sound attributes. Recent decoding studies have shown that parietal cortex supports the storage of content-specific representations (Christophel et al., 2012; Ester et al., 2015; Bettencourt and Xu, 2016; Galeano Weber et al., 2016; Xu, 2017; Lorenc et al., 2018; Rademaker et al., 2019) and supports content-specific prioritization of memory representations (Peters et al., 2015). The present decoding of location and pitch thus supports an overarching role of superior parietal cortex also for auditory WM. The higher decoding accuracy for location than pitch could reflect the fact that the selected ROI combined subregions underlying WM storage and attentional prioritization on the one hand and those involved in the processing of spatial information in the visual and auditory domains on the other hand, which has been associated with parietal cortex (Xu, 2018).

As persistent delay-period activity in prefrontal cortex traditionally has been considered the neural basis of WM storage, we also included a middle frontal ROI in our analyses. There has been a debate about the function of lateral prefrontal cortex in WM (Sreenivasan et al., 2014), with more recent work supporting the decodability of visual WM contents in this region (for review, see Serences, 2016; Christophel et al., 2017). In contrast to the higher-order superior parietal ROI, neither location nor pitch could be decoded from this region. As visual studies have found decoding of memorized stimuli in relatively small subregions of lateral frontal cortex only (Ester et al., 2015), it is possible that comparable representations of auditory stimuli are located dorsally or ventrally (Kumar et al., 2016; Uluç et al., 2018) of the present middle frontal ROI. Indeed, we found comparable decoding accuracy for location and pitch information in the neighboring inferior frontal ROI.

Our finding of concurrent decodability of different object features in superior parietal and, to a lesser extent, inferior frontal cortex might also contribute to the debate about the format of

information held in WM, i.e., whether objects in WM are stored as independent features or as feature bindings (Schneegans and Bays, 2017). Previous research on auditory perception showing processing of multiple object features in distinct anatomic pathways has pointed toward the idea of independent feature stores (Arnott et al., 2004). In contrast, a recent MVPA study has suggested a role for posterior parietal cortex in the integration of sound frequency and amplitude modulation (Sohoglu et al., 2020). Our study suggests that multifeature memory information is represented concurrently within the same anatomic brain regions including higher-order frontal and parietal cortex, opening the possibility for coding in a feature binding format, e.g., via a binding pool or binding space (Swan and Wyble, 2014; Oberauer and Lin, 2017).

The present MVPA (uncorrected for multiple comparisons across ROIs) resulted in a classification performance that was lower than for decoding of visual WM stimuli (Harrison and Tong, 2009; Emrich et al., 2013), but comparable with levels reported by other auditory WM studies (Kumar et al., 2016). This discrepancy may be attributable to the fact that fMRI investigations of auditory processing are inevitably affected by the noise produced by the MR scanner, even when using an active noise cancellation system as in the present investigation. Same-modality distractors presented during the delay interval have been shown to reduce classification performance particularly in sensory cortex (Bettencourt and Xu, 2016; Rademaker et al., 2019). Moreover, a comparison of our results with studies requiring the maintenance of only one single-feature item is problematic, as the maintenance of two stimulus features increased memory demands, which is known to lead to decreased decoding accuracy (Emrich et al., 2013; Sprague et al., 2014; Gosseries et al., 2018). Our results showed that multiple features of a single sound object were decodable in sensory and higher-order cortical regions despite the distracting auditory noise produced by the MR scanner and despite the increased memory demands associated with retaining two task-relevant features concurrently in WM.

References

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) "What" and "where" in the human auditory system. Proc Natl Acad Sci USA 98:12301–12306.

Alain C, He Y, Grady C (2008) The contribution of the inferior parietal lobe to auditory spatial working memory. J Cogn Neurosci 20:285–295.

Alain C, Khatamian Y, He Y, Lee Y, Moreno S, Leung AWS, Bialystok E (2018) Different neural activities support auditory working memory in musicians and bilinguals. Ann NY Acad Sci 1423:435–446.

Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20:870–888.

Arnott SR, Binns MA, Grady CL, Alain C (2004) Assessing the auditory dual-pathway model in humans. Neuroimage 22:401–408.

Arnott SR, Grady CL, Hevenor SJ, Graham S, Alain C (2005) The functional organization of auditory working memory as revealed by fMRI. J Cogn Neurosci 17:819–831.

Backer KC, Binns MA, Alain C (2015) Neural dynamics underlying attentional orienting to auditory representations in short-term memory. J Neurosci 35:1307–1318.

Baddeley A (1992) Working memory. Science 255:556–559.

Bettencourt KC, Xu Y (2016) Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nat Neurosci 19:150–157.

Bledowski C, Rahm B, Rowe JB (2009) What "works" in working memory? Separate systems for selection and updating of critical information. J Neurosci 29:13735–13741.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Christophel TB, Hebart MN, Haynes JD (2012) Decoding the contents of visual short-term memory from human visual and parietal cortex. J Neurosci 32:12983–12989.

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes JD (2017) The distributed nature of working memory. Trends Cogn Sci 21:111–124.

Christophel TB, Iamshchinina P, Yan C, Allefeld C, Haynes JD (2018) Cortical specialization for attended versus unattended working memory. Nat Neurosci 21:494–496.

Cichy RM, Chen Y, Haynes JD (2011) Encoding the identity and location of objects in human LOC. Neuroimage 54:2297–2307.

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386:608–611.

Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci 7:415–423.

D'Esposito M, Postle BR (2015) The cognitive neuroscience of working memory. Annu Rev Psychol 66:115–142.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980.

Emrich SM, Riggall AC, Larocque JJ, Postle BR (2013) Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J Neurosci 33:6516–6523.

Erhart M, Czoschke S, Fischer C, Bledowski C, Kaiser J (2021) Decoding spatial versus non-spatial processing in auditory working memory. Front Neurosci 15:637877.

Ester EF, Sprague TC, Serences JT (2015) Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. Neuron 87:893–905.

Fischer C, Czoschke S, Peters B, Rahm B, Kaiser J, Bledowski C (2020) Context information supports serial dependence of multiple visual objects across memory episodes. Nat Commun 11:1932.

Foster JJ, Bsales EM, Jaffe RJ, Awh E (2017) Alpha-band activity reveals spontaneous representations of spatial position in visual working memory. Curr Biol 27:3216–3223.e6.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Gaab N, Gaser C, Zaehle T, Jäncke L, Schlaug G (2003) Functional anatomy of pitch memory–an fMRI study with sparse temporal sampling. Neuroimage 19:1417–1426.

Galeano Weber EM, Peters B, Hahn T, Bledowski C, Fiebach CJ (2016) Superior intraparietal sulcus controls the variability of visual working memory precision. J Neurosci 36:5623–5635.

Gosseries O, Yu Q, LaRocque JJ, Starrett MJ, Rose NS, Cowan N, Postle BR (2018) Parietal-occipital interactions underlying control- and representation-related processes in working memory for nonspatial visual features. J Neurosci 38:4357–4366.

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635.

Haynes JD (2015) A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. Neuron 87:257–270.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci USA 103:3863–3868.

Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. Cognition 80:97–126.

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.

LaRocque JJ, Riggall AC, Emrich SM, Postle BR (2017) Within-category decoding of information in different attentional states in short-term memory. Cereb Cortex 27:4881–4890.

Leung AW, Alain C (2011) Working memory load modulates the auditory "what" and "where" neural networks. Neuroimage 55:1260–1269.

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.

Lim SJ, Wöstmann M, Obleser J (2015) Selective attention to auditory memory neurally enhances perceptual precision. J Neurosci 35:16094–16104.

Linke AC, Vicente-Grabovetsky A, Cusack R (2011) Stimulus-specific suppression preserves information in auditory short-term memory. Proc Natl Acad Sci USA 108:12961–12966.

Long NM, Kuhl BA (2018) Bottom-up and top-down factors differentially influence stimulus representations across large-scale attentional networks. J Neurosci 38:2495–2504.

Lorenc ES, Sreenivasan KK, Nee DE, Vandenbroucke ARE, D'Esposito M (2018) Flexible coding of visual working memory representations during distraction. J Neurosci 38:5267–5276.

Ma WJ, Husain M, Bays PM (2014) Changing concepts of working memory. Nat Neurosci 17:347–356.

Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643.

Nobre AC, Coull JT, Maquet P, Frith CD, Vandenberghe R, Mesulam MM (2004) Orienting attention to locations in perceptual versus mental representations. J Cogn Neurosci 16:363–373.

Oberauer K, Lin HY (2017) An interference model of visual working memory. Psychol Rev 124:21–59.

Pessoa L, Gutierrez E, Bandettini P, Ungerleider L (2002) Neural correlates of visual working memory: fMRI amplitude predicts task performance. Neuron 35:975–987.

Peters B, Kaiser J, Rahm B, Bledowski C (2015) Activity in human visual and parietal cortex reveals object-based attention in working memory. J Neurosci 35:3360–3369.

Petrides M (2000) Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. J Neurosci 20:7496–7503.

Rademaker RL, Chunharas C, Serences JT (2019) Coexisting representations of sensory and mnemonic information in human visual cortex. Nat Neurosci 22:1336–1344.

Rämä P, Poremba A, Sala JB, Yee L, Malloy M, Mishkin M, Courtney SM (2004) Dissociable functional cortical topographies for working memory maintenance of voice identity and location. Cereb Cortex 14:768–780.

Rauschecker JP (1998) Parallel processing in the auditory cortex of primates. Audiol Neurootol 3:86–103.

Riggall AC, Postle BR (2012) The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J Neurosci 32:12990–12998.

Rinne T, Koistinen S, Salonen O, Alho K (2009) Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. J Neurosci 29:13338–13343.

Rinne T, Koistinen S, Talja S, Wikman P, Salonen O (2012) Task-dependent activations of human auditory cortex during spatial discrimination and spatial memory tasks. Neuroimage 59:4126–4131.

Schneegans S, Bays PM (2017) Neural architecture for feature binding in visual working memory. J Neurosci 37:3913–3925.

Serences JT (2016) Neural mechanisms of information storage in visual short-term memory. Vision Res 128:53–67.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207–214.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 [Suppl 1]:S208–S219.

Sohoglu E, Kumar S, Chait M, Griffiths TD (2020) Multivoxel codes for representing and integrating acoustic features in human cortex. Neuroimage 217:116661.

Sprague TC, Ester EF, Serences JT (2014) Reconstructions of information in visual spatial working memory degrade with memory load. Curr Biol 24:2174–2180.

Sreenivasan KK, Curtis CE, D'Esposito M (2014) Revisiting the role of persistent neural activity during working memory. Trends Cogn Sci 18:82–89.

Swan G, Wyble B (2014) The binding pool: a model of shared neural resources for distinct items in visual working memory. Atten Percept Psychophys 76:2136–2157.

Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428:751–754.

Treder MS (2020) MVPA-light: a classification and regression toolbox for multi-dimensional data. Front Neurosci 14:289.

Uluç I, Schmidt TT, Wu YH, Blankenburg F (2018) Content-specific codes of parametric auditory working memory in humans. Neuroimage 183:254–262.

van Ede F, Chekroud SR, Nobre AC (2019) Human gaze tracks attentional focusing in memorized visual space. Nat Hum Behav 3:462–470.

van Loon AM, Olmos-Solis K, Fahrenfort JJ, Olivers CN (2018) Current and future goals are represented in opposite patterns in object-selective cortex. Elife 7:e38677.

Wolff MJ, Jochim J, Akyürek EG, Stokes MG (2017) Dynamic hidden states underlying working-memory-guided behavior. Nat Neurosci 20:864–871.

Xu Y (2017) Reevaluating the sensory account of visual working memory storage. Trends Cogn Sci 21:794–815.

Xu Y (2018) The posterior parietal cortex in adaptive visual processing. Trends Neurosci 41:806–822.

Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.