

**Title: Unimodal and bimodal access to sensory working memories
by auditory and visual impulses**

Abbreviated title: Unimodal and bimodal access to sensory WM content

Authors: Wolff, M. J.^{1,2}, Kandemir, G.¹, Stokes, M. G.², & Akyürek, E. G.¹

1. Department Experimental Psychology, University of Groningen, Groningen, 9712 TS,
The Netherlands
2. Department Experimental Psychology, University of Oxford, Oxford, OX2 6GG,
United Kingdom

Corresponding author: Michael J. Wolff, michael.wolff@psy.ox.ac.uk

Number of pages: 43

Number of figures: 6

Number of words: 208 (Abstract), 606 (Introduction), 1497 (Discussion)

Conflict of interest statement

The authors declare no competing financial interests.

Acknowledgments

This research was in part funded by an ESRC (ES/S015477/1) grant and James S. McDonnell Foundation Scholar Award (220020405) to MGS, and by the NIHR Oxford Health Biomedical Research Centre. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health. We would like to thank P.

- 22 Albronda for providing technical support and M. Rietdijk for helping with data collection, as
23 well as N.E. Myers and S. Hall-McMaster for helpful discussion.

24

Abstract

25 It is unclear to what extent sensory processing areas are involved in the maintenance of
26 sensory information in working memory (WM). Previous studies have thus far relied on
27 finding neural activity in the corresponding sensory cortices, neglecting potential activity-
28 silent mechanisms such as connectivity-dependent encoding. It has recently been found that
29 visual stimulation during visual WM maintenance reveals WM-dependent changes through a
30 bottom-up neural response. Here, we test whether this impulse response is uniquely visual
31 and sensory-specific. Human participants (both sexes) completed visual and auditory WM
32 tasks while electroencephalography was recorded. During the maintenance period, the WM
33 network was perturbed serially with fixed and task-neutral auditory and visual stimuli. We
34 show that a neutral auditory impulse-stimulus presented during the maintenance of a pure
35 tone resulted in a WM-dependent neural response, providing evidence for the auditory
36 counterpart to the visual WM findings reported previously. Interestingly, visual stimulation
37 also resulted in an auditory WM-dependent impulse response, implicating the visual cortex in
38 the maintenance of auditory information, either directly, or indirectly as a pathway to the
39 neural auditory WM representations elsewhere. In contrast, during visual WM maintenance
40 only the impulse response to visual stimulation was content-specific, suggesting that visual
41 information is maintained in a sensory-specific neural network, separated from auditory
42 processing areas.

43

Significance Statement

44 Working memory is a crucial component of intelligent, adaptive behaviour. Our
45 understanding of the neural mechanisms that support it has recently shifted: rather than being
46 dependent on an unbroken chain of neural activity, working memory may rely on transient
47 changes in neuronal connectivity, which can be maintained efficiently in activity-silent brain

48 states. Previous work using a visual impulse stimulus to perturb the memory network has
49 implicated such silent states in the retention of line orientations in visual working memory.
50 Here, we show that auditory working memory similarly retains auditory information. We also
51 observed a sensory-specific impulse response in visual working memory, while auditory
52 memory responded bi-modally to both visual and auditory impulses, possibly reflecting
53 visual dominance of working memory.

Introduction

Working memory (WM) is necessary to maintain information without sensory input, which is vital to adaptive behaviour. In spite of its important role, it is not yet fully clear how WM content is represented in the brain, or whether sensory information is maintained within a sensory-specific neural network. Previous research has relied on testing whether sensory cortices exhibit content-specific neural activity during maintenance. While this has indeed been shown for visual memories in occipital areas (e.g., Harrison & Tong, 2009) and, more recently, for auditory memories in the auditory cortex (Huang, Matysiak, Heil, König, & Brosch, 2016; Kumar et al., 2016; Uluç, Schmidt, Wu, & Blankenburg, 2018), WM-specific activity in the sensory cortex is not always present (Bettencourt & Xu, 2016), fuelling an ongoing debate over whether sensory cortices are necessary for WM maintenance (Scimeca, Kiyonaga, & D'Esposito, 2018; Xu, 2017). However, the neural WM network may not be solely based on measurable neural activity, and it has been proposed that information in WM may be maintained in an “activity-silent” network (Stokes, 2015) – for example, through changes in short-term connectivity (Mongillo, Barak, & Tsodyks, 2008). Potentially silent WM states should also be taken into account to better investigate the sensory-specificity account of WM.

Silent network theories predict that its neural impulse response to external stimulation can be used to infer its current state (Buonomano & Maass, 2009; Stokes, 2015). This has been shown in visual WM experiments, in which the evoked neural response from a fixed, neutral and task-irrelevant visual stimulus presented during the maintenance period of a visual WM task contained information about the contents of visual WM (Wolff, Ding, Myers, & Stokes, 2015; Wolff, Jochim, Akyürek, & Stokes, 2017). This not only suggests that otherwise hidden processes can be illuminated, but also implicates the involvement of the visual cortex in the maintenance of visual information, even when no ongoing activity can be detected. It

has been suggested that this WM-dependent response profile might be not merely a by-product of connectivity-dependent WM, but a fundamental mechanism that affords efficient and automatic readout of WM content through external stimulation (Myers et al., 2015). It remains an open question, however, whether information from other modalities in WM is similarly organized. If auditory WM depends on content-specific connectivity changes that include the sensory cortex, we would expect a network-specific neural response to external auditory stimulation. Furthermore, it may be hypothesized that sensory information need not necessarily be maintained in a network that is detached from other sensory processing areas. Direct connectivity (Eckert et al., 2008) and interplay (Iurilli et al., 2012; Martuzzi et al., 2007) between the auditory and visual cortices, or areas where information from different modalities converges, such as the parietal and pre-frontal cortices (Driver & Spence, 1998; Stokes et al., 2013), raise the possibility that WM could exploit these connections even during maintenance of unimodal information. Content-specific impulse responses might be observed not only during sensory-specific but also sensory non-specific stimulation.

In the present study, we tested whether WM-dependent impulse responses can be observed in visual and auditory WM, and whether that response is sensory specific. We measured electroencephalography (EEG) while participants performed visual and auditory WM tasks. We show that the evoked neural response of an auditory impulse stimulus reflects relevant auditory information maintained in WM. Visual perturbation also resulted in an auditory WM-dependent neural response, implicating both the auditory and visual cortices in auditory WM. By contrast, visual WM content could only be decoded after visual, but not auditory perturbation, suggesting that visual information is maintained in a sensory-specific visual WM network with no evidence for a WM-related interplay with the auditory cortex.

Materials and Methods

Participants

Thirty healthy adults (12 female, mean age 21 years, range 18-31 years) were included in the main analyses of the auditory WM experiment and 28 healthy adults (11 female, mean age 21 years, range 19-31 years) of the visual WM experiment. Three additional participants in the auditory WM experiment and 8 additional participants in the visual WM experiment were excluded during pre-processing due to excessive eye movements (more than 30% of impulse epochs contaminated). The exclusion criterion and resulting minimum number of trials for the multivariate pattern analysis were similar to our previous study (Wolff et al., 2017). Participants received either course credits or monetary compensation (8€ an hour) for participation and gave written informed consent. Both experiments were approved by the Departmental Ethical Committee of the University of Groningen (approval number: 16109-S-NE).

Apparatus and Stimuli

Stimuli were controlled by Psychtoolbox, a freely available toolbox for Matlab. Visual stimuli were generated with Psychtoolbox and presented on a 17-inch (43.18 cm) CRT screen running at 100 Hz refresh rate and a resolution of 1280 by 1024 pixels. Auditory stimuli were generated with the freely available software Audacity and were presented with stereo Logitech computer speakers. The intensity of all tones was adjusted to 70 dB SPL at a fixed distance of 60 cm between speakers and participants in both experiments. All tones had 10 ms ramp up and ramp down time. Responses were collected with a custom two-button response box, connected via a USB interface.

The memory items used in the auditory WM experiment were 8 pure tones, ranging from 270 Hz to 3055 Hz in steps of half an octave. The probes in the auditory experiment were 16 pure

tones that were one-third of an octave higher or lower than the corresponding auditory memory items.

The memory items used in the visual WM experiment were 8 sine-wave gratings with orientations of 11.25° to 168.75° in steps of 22.5°. The visual probes were 16 sine-wave gratings that were rotated 20° clockwise or counter-clockwise relative to the corresponding visual memory items. All gratings were presented at 20% contrast, with a diameter of 6.5° (at 60 cm distance) and a spatial frequency of 1 cycle per degree. The phase of each grating was randomized within and across trials.

The remaining stimuli were the same in both experiments. The retro-cue was a number (“1” or “2”) that subtended 0.7°. The visual impulse stimulus was a white circle with a diameter of 12°. The auditory impulse was a complex tone consisting of the combination of all pure tones used as memory items in the auditory task. A grey background (RGB = 128, 128, 128) and a black fixation dot with a white outline (0.25°) were maintained throughout the trials. All visual stimuli were presented in the centre of the screen.

Experimental Design

The trial structure was the same in both experiments, as shown in Figure 1 (panels A and C). In both cases, participants completed a retro-cue WM task. Only the memory items and probes differed between experiments. Memory items and probes were pure tones in the auditory WM task and sine-wave gratings in the visual WM task. Each trial began with the presentation of a fixation dot, which stayed on the screen throughout the trial. After 1,000 ms the first memory item was presented for 200 ms. After a 700 ms delay the second memory item in the same modality as the first item was presented for 200 ms. Each memory item was selected randomly without replacement from a uniform distribution of 8 different tonal frequencies or grating orientations (see above) for the auditory and visual experiment,

respectively. After another delay of 700 ms, the retro-cue was presented for 200 ms, indicating to participants whether the first or second memory item would be tested at the end of the trial. After a delay of 1,000 ms the impulse stimuli (the visual circle and the complex tone) were presented serially for 100 ms each with a delay of 900 ms in-between. The order of the impulses was fixed for each participant but counter-balanced between participants. Impulse order was fixed within participants for two reasons: First, it removed the effect of surprise by making the order of events within trials perfectly consistent and predictable (Wessel & Aron, 2017), ensuring minimal intrusion by the impulse stimuli during the maintenance period. Second, random impulse order might have resulted in qualitatively different neural responses of each impulse, depending on when it was presented, due to different trial histories and elapsed maintenance duration at the time of impulse onset (Buonomano & Maass, 2009). This would have necessitated splitting the neural data by impulse order for the decoding analyses, resulting in reduced power. The probe stimulus followed 900 ms after the second impulse offset and was presented for 200 ms. In the auditory WM experiment the probe was a pure tone and the participant's task was to indicate via button press on the response box whether the probe's frequency was lower (left button) or higher (right button) than the cued memory item. In the visual task, the probe was another visual grating, and the participants indicated whether it was rotated counter-clockwise (left button) or clockwise (right button) relative to the cued memory item. The direction of the tone or tilt was selected randomly without replacement from a uniform distribution. After each response, a smiley face was shown for 200 ms, which indicated whether the response was correct or incorrect. The next trial began automatically after a randomized, variable delay of 700-1,000 ms after response input. Each experiment consisted of 768 trials in total and lasted approximately 2 hours.

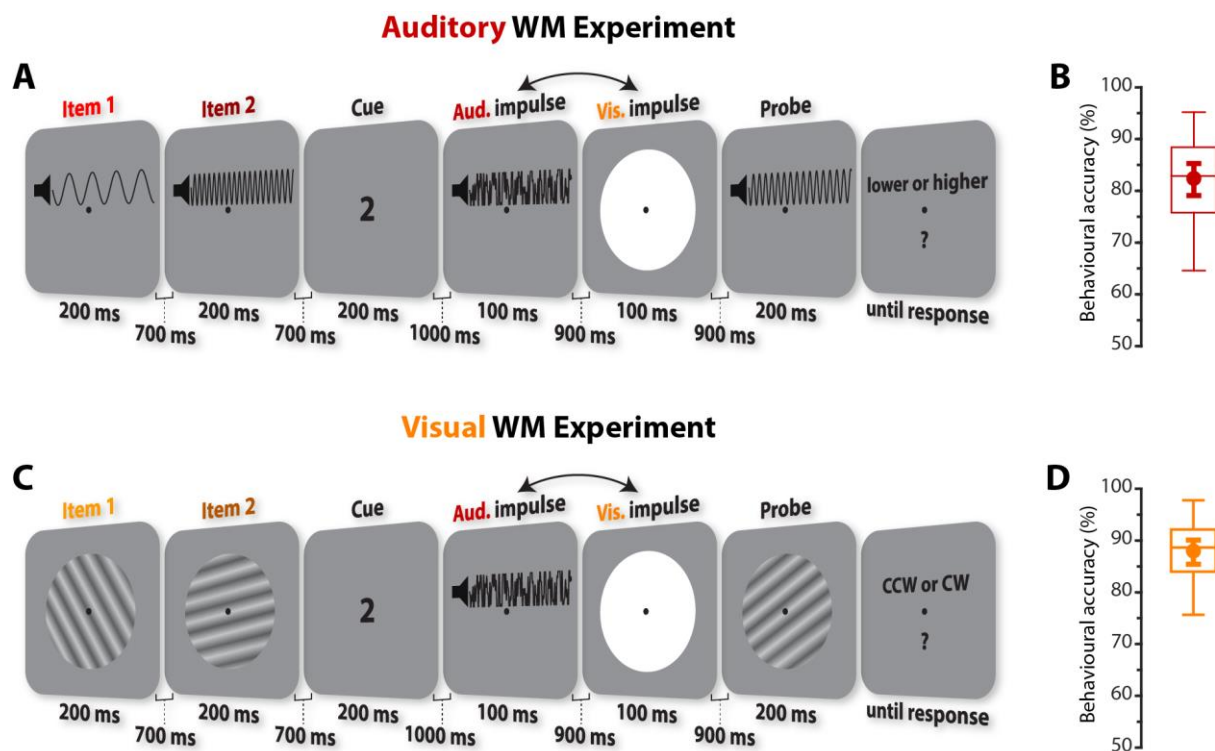


Figure 1. Task structure and behavioural performance

(A) Trial schematic of auditory task. Two randomly selected pure tones (270 Hz to 3055 Hz) were serially presented and a retro-cue indicated which of those tones would be tested at the end of the trial. In the subsequent delay, two irrelevant impulse stimuli (a complex tone and a white circle) were serially presented. At the end of each trial another pure tone was presented (the probe), and participants were instructed to indicate whether the frequency of the previously cued tone was higher or lower than the probe's frequency. (B) The boxplot shows auditory task accuracy. Centre line indicates the median; box outlines show 25th and 75th percentiles, and whiskers indicate 1.5x the interquartile range. The superimposed circle and error bars indicate the mean and its 95% C.I., respectively. (C) Trial schematic of visual task. The trial structure was the same as in the auditory task. Instead of pure tones, memory items were randomly orientated gratings. The probe was another orientation grating, and participants were instructed to indicate whether the cued item's orientation was rotated

clockwise or counter-clockwise relative to the probe's orientation. **(D)** Visual task performance.

EEG acquisition and pre-processing

The EEG signal was acquired from 62 Ag/AgCl sintered electrodes laid out according to the extended international 10-20 system. An analog-to-digital TMSI Refa 8-64/72 amplifier and Brainvision recorder software were used to record the data at 1000 Hz using an online average reference. An electrode placed just above the sternum was used as the ground. Bipolar electrooculography (EOG) was recorded by electrodes placed above and below the right eye, and to the left and right of the left and right eye, respectively. The impedances of all electrodes were kept below 10 k Ω .

Offline the data was down-sampled to 500 Hz and bandpass filtered (0.1 Hz high-pass and 40 Hz low-pass) using EEGLAB (Delorme & Makeig, 2004). The data was epoched relative to the onsets of the memory items (-150 ms to 900 ms) and to the onsets of the auditory and visual impulse stimuli (-150 to 500 ms). The signal's variance across channels and trials was visually inspected using a visualization tool provided by the Matlab extension Fieldtrip (Oostenveld et al., 2010), and especially noisy channels were removed and replaced through spherical interpolation. This led to the interpolation of 1 channel in 3 participants and 2 channels in 1 participant in the auditory WM task, and 1 channel in 5 participants and 5 channels in 1 participant in the visual WM task. Noisy epochs were removed from all subsequent electrophysiological analyses. Epochs containing any artefacts related to eye movements were identified by visually inspecting the EOG signals and also removed from analyses. The following percentage of trials were removed for each epoch in the auditory WM experiment: item 1 epoch ($M = 13.39\%$, $SD = 6.08\%$), item 2 epoch ($M = 9.28\%$, $SD =$

4.42%), auditory impulse epoch ($M = 11.53\%$, $SD = 7.03\%$), visual impulse epoch ($M = 9.81\%$, $SD = 5.44\%$). The following percentage of trials were removed for each epoch in the visual WM experiment: item 1 epoch ($M = 19.81\%$, $SD = 5.91\%$), item 2 epoch ($M = 20.69\%$, $SD = 5.88\%$), auditory impulse epoch ($M = 18.51\%$, $SD = 5.73\%$), visual impulse epoch ($M = 19.33\%$, $SD = 4.94\%$).

Multivariate pattern analysis of neural dynamics

We wanted to test if the electrophysiological activity evoked by the memory stimuli and impulse stimuli contained item-specific information. Since event-related potentials are highly dynamic, we used an approach that is sensitive to such changing neural activity within pre-defined time-windows, by pooling relative voltage fluctuations over space (i.e., electrodes) and time. This approach has two key benefits: First, pooling information over time (in addition to space) multivariately can boost decoding accuracy (Grootswagers, Wardle, & Carlson, 2017; Nemrodov, Niemeier, Patel, & Nestor, 2018). Secondly, by removing the mean-activity level within each time-window, the voltage fluctuations are normalized. This is similar to taking a neutral pre-stimulus baseline, as is common in ERP analysis. Notably, this also removes stable activity traces that do not change within the chosen time-window, making this approach ideal to decode transient, stimulus-evoked activation patterns, while disregarding more stationary neural processes. The following details of the analyses were the same for each experiment, unless explicitly stated.

For the time-course analysis, we used a sliding window approach that takes into account the relative voltage changes within a 100 ms window. The time-points within 100 ms of each channel and trial were first down-sampled by taking the average every 10 ms, resulting in 10 voltage values for each channel. Next, the mean activity within that time-window of each

channel was subtracted from each individual voltage value. All 10 voltage values per channel were then used as features for the 8-fold cross-validation decoding approach.

We used Mahalanobis distance (De Maesschalck, Jouan-Rimbaud, & Massart, 2000) to take advantage of the potentially parametric neural activity underlying the processing and maintenance of orientations and tones. The distances between each of the left-out test-trials and the averaged, condition-specific patterns of the train-trials (tones and orientations in the auditory and visual experiment, respectively), were computed, with the covariance matrix estimated from the train-trials using a shrinkage estimator (Ledoit & Wolf, 2004). To acquire reliable distance estimates, this process was repeated 50 times, where the data was randomly partitioned into 8 folds using stratified sampling each time. The number of trials of each condition (orientation/tone frequency) of the 7 train-folds were equalized by randomly subsampling the minimum number of condition-specific trials to ensure an unbiased training set. The average was then taken of these repetitions. For each trial the 8 distances (one of each stimulus condition) were sign-reversed for interpretation purposes, so that higher values reflect higher pattern-similarity between test and train-trials. For visualization, the sign-reversed distances were furthermore mean-centred by subtracting the mean distance of all distances of a given trial and ordered as a function of tone difference, in 1 octave steps by averaging over adjacent half octave differences, and orientation differences.

To summarize the expected positive relationship between tone-similarity and neural activation similarity (indicative of tone-specific information in the recorded signal) into a single value in the auditory WM experiment, the absolute tonal differences were linearly regressed against the corresponding pattern similarity values for each trials. The obtained beta values of the slopes were then averaged across all trials to represent “decoding accuracy”, where high values suggest a strong positive effect of tone similarity on neural pattern similarity. To summarize the tuning curves in the visual WM experiment, we

computed the cosine vector means (Wolff et al., 2017), where high values suggest evidence for orientation decoding.

The approach described above was repeated in steps of 8 ms across time (-52 to 900 ms relative to item 1 and 2 onset, and -52 to 500 ms relative to auditory and visual onset). The decoding values were averaged over trials, and the decoding time-course was smoothed with a Gaussian smoothing kernel (s.d. = 16 ms). Within the time-window, information was pooled from -100 to 0 ms relative to a specific time-point. By only including data-points from before the time-point of interest, it is ensured that decoding onsets can be more easily interpreted, whereas decoding offsets should be interpreted with caution (Grootswagers et al., 2017). In addition to the sliding window approach, we also pooled information multivariately across the whole time-window of interest (Nemrodov et al., 2018). As before, the data was first down-sampled by taking the average every 10 ms, and the mean activity from 100 to 400 ms relative to impulse onset was subtracted. The resulting 30 values per channel were then provided to the multivariate decoding approach in the same way as above, resulting in a single decoding value per participant. The time-window of interest was based on previous findings showing that the WM-dependent impulse response is largely confined within that window (Wolff et al., 2017). Additionally, items in the item-presentation epochs were also decoded using each channel separately, using the data from 100-400 ms relative to onset. Decoding topographies were visualized using fieldtrip (Oostenveld et al., 2010).

Cross-epoch generalization analysis

We also tested if WM-related decoding in the impulse epochs generalized to the memory presentation. Instead of using the same epoch (100-400 ms) for training and testing, as described above, the classifier was trained on the memory item epoch and tested on the impulse epoch that contained significant item decoding (and vice versa). In the auditory task,

we also tested if the different impulse epochs cross-generalized by training on the visual and testing on the auditory impulse (and vice versa).

Representational similarity analysis

While the decoding approach outlined above takes into account the potentially parametric relationship of pitch/orientation difference, it is not an explicit test for the presence of a parametric relationship. Indeed, decodability could theoretically be solely driven by high within stimulus-condition pattern similarity, and equally low pattern similarities of all between stimulus-condition comparisons. To explicitly test for a linear/circular relationship between stimuli, and explore additional stimulus coding schemes, we used representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008).

The RSA was based on the mahalanobis distances between all stimulus conditions (unique orientations and frequencies) in both experiments using the same time-window of interest as in the decoding approach described above (100-400 ms relative to stimulus onset). For each participant, the number of trials of each stimulus condition were equalized by randomly subsampling the minimum number of trials of a condition before taking the average across all same stimulus condition trials and computing all pairwise mahalanobis distances. This procedure was repeated 50 times, with random subsamples each time, before averaging them all into a single representation dissimilarity matrix (RDM). The covariance matrix was computed from all trials using the shrinkage estimator (Ledoit & Wolf, 2004). Since each experiment contained 8 unique memory items, this resulted in an 8 x 8 RDM for each participant and epoch of interest.

For the RSA in the auditory WM experiment we considered two models; a positive linear relationship between absolute pitch height difference (i.e., the more dissimilar pitch frequency, the more dissimilar the brain activity patterns), and a positive relationship of pitch

chroma (i.e. higher similarity between brain activity patterns of the same pitch chromas). Note that the tone frequencies used in the experiment increased in half octave steps. Every other tone had thus the same pitch chroma (i.e. the same note in a different octave). The model RDMs are shown for illustration in Figure 4A. The model RDMs were z-scored to make the corresponding model fits between them more comparable, before entering both of them into a multiple regression analysis with the data RDM.

In the visual WM experiment we also considered two models. The first model was designed to capture the circular relationship between absolute orientation difference (i.e., the more dissimilar the orientation, the more dissimilar the brain activity patterns). The second model was designed to capture the specialization of cardinal orientations (i.e., horizontal and vertical) that could reflect the “oblique effect”, where orientations close to the cardinal axes are discriminated and recalled more accurately than more oblique orientations (Appelle, 1972; Pratte, Park, Rademaker, & Tong, 2017). The model assumed the extreme case, where orientations are clustered into one of three categories depending on their circular distance to vertical, horizontal, or oblique angles. This captures the relatively higher dissimilarity and distinctiveness of the cardinal axes (vertical and horizontal) compared to the oblique axes (-45 degrees and +45 degrees) and reflects neurophysiological findings of an increased number of neurons tuned to the cardinal axes (Shen, Tao, Zhang, Smith, & Chino, 2014). The model RDMs are shown for illustration in Figure 4D. The model RDMs were also z-scored and then both included into a multiple regression with the data RDM.

Statistical Analysis

All statistical tests were the same between experiments. Sample sizes of all analyses were $n=30$ and $n=28$ in the auditory and visual tasks, respectively. Sample size of the event-related potential (ERP) analyses as a function of impulse modality and task was $n=16$, as it only

included participants who participated in both WM tasks. To determine if the decoding values (see above) or model fits of the RSA are higher than 0 or different between items, or if the evoked potentials were different between tasks, we used a non-parametric sign-permutation test (Maris & Oostenveld, 2007). The sign of the decoding value, model fit value, or voltage difference of each participant was randomly flipped 100.000 times with a probability of 50%. The p value was derived from the resulting null distribution. The above procedure was repeated for each time-point for time-series results. A cluster based permutation test (100.000 permutations) was used to correct for multiple comparisons over time using a cluster forming and cluster significance threshold of $p < 0.05$. Complementary Bayes factors to test for decoding evidence for the cued and uncued items within each impulse epoch separately were also computed.

We were also interested if there were differential effects on the decoding results between cueing (cued/uncued) and impulse modality (auditory/visual) during WM maintenance. To test this, we computed the Bayes factors of models with and without each of these predictors versus the null model that only included subjects as a predictor (Bayesian equivalent of repeated measures ANOVA). The freely available software package JASP (JASP Team, 2018) was used to compute Bayes factors.

Differences in behavioural performance between tasks were tested with the partially overlapping samples t-test (Derrick, Toher, & White, 2017), since only some participants took part in both tasks. No violations of normality or equality of variances were detected.

Code and data availability

All data and custom Matlab scripts used to generate the results and figures of this manuscript are available from the OSF database ([osf.io/ u7k3q](https://osf.io/u7k3q)).

Results

Behavioural results

Behavioural task performance was $M = 82.322\%$, $SD = 8.841\%$ in the auditory WM task (Fig. 1B), and $M = 87.908\%$, $SD = 6.374\%$ in the visual WM task (Fig. 1D). Performance was significantly higher in the visual than in the auditory task, $t(33.379) = 2.776$, $p = 0.009$, two-sided. Despite this difference, it is clear that participants performed well above chance in both tasks, suggesting that the relevant sensory features were reliably remembered and recalled.

Decoding visual and auditory stimuli

Auditory WM task

The neural dynamics of auditory stimulus processing suggest a parametric effect, with a positive relationship between tone and pattern similarity (Fig. 2A & C) for both memory items. The neural dynamics showed significant item-specific decoding clusters during, and shortly after, corresponding item presentation for item 1 (44 to 708 ms relative to item 1 onset, $p < 0.001$, one-sided, corrected) and item 2 (28 to 572 ms relative to item 2 onset, $p < 0.001$, one-sided, corrected; Fig. 2B). The topographies of channel-wise item decoding for each item using the neural data from 100-400 ms after item onset, revealed strong decoding for frontal-central and lateral electrodes (Fig. 2D), suggesting that the tone-specific neural activity is most likely generated by the auditory cortex (Chang, Bosnyak, & Trainor, 2016). These results provide evidence that stimulus-evoked neural activity fluctuations contain information about presented tones that can be decoded from EEG.

Visual WM task

Processing of visual orientations also showed a parametric effect (Fig. 2E & G), replicating previous findings (Saproo & Serences, 2010). The item-specific decoding time-courses of the dynamic activity showed significant decoding clusters during and shortly after item presentation (item 1: 84-724 ms, $p < 0.001$; item 2: 84-636 ms, $p < 0.001$, one-sided, corrected). As expected, the topographies of channel-wise item-decoding showed strong effects in posterior channels that are associated with the visual cortex.

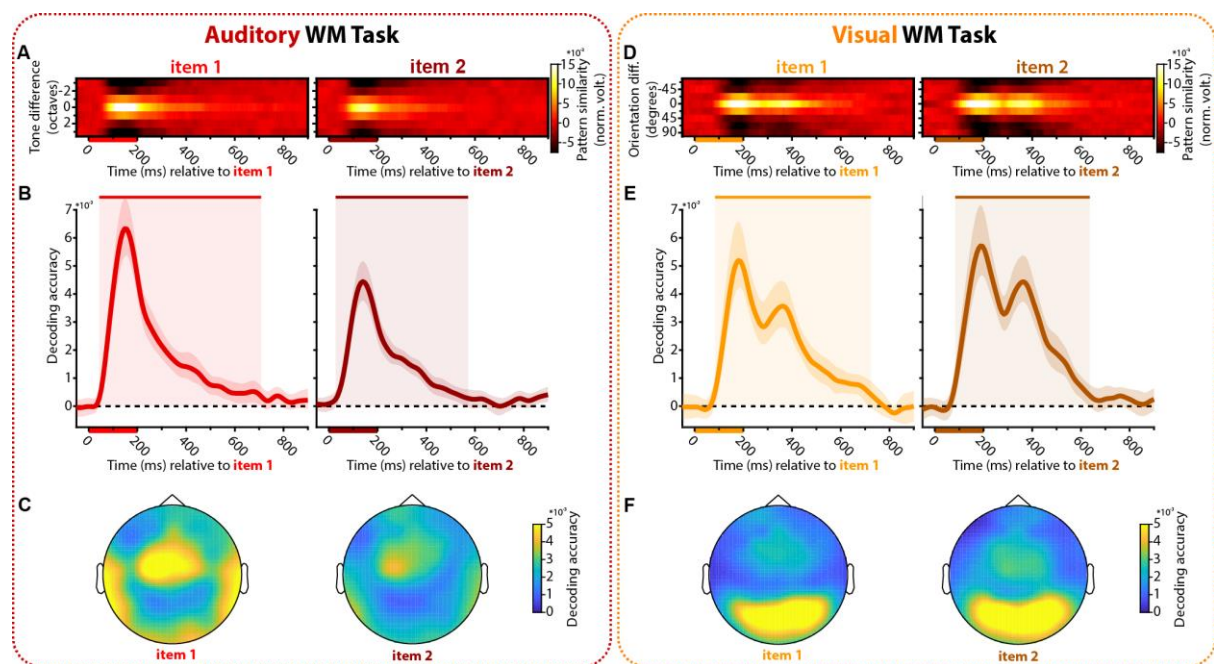


Figure 2. Decoding during item encoding

A-C, Auditory WM task. **D-F**, Visual WM task. **A & D**, Normalized average pattern similarity (mean-centred, sign-reversed mahalanobis distance) of the neural dynamics for each time-point between trials as a function of tone similarity in **A** and orientation similarity in **D**, separately for item 1 and item 2, in item 1 and item 2 epochs, respectively. Bars on the horizontal axes indicate item presentations. **B & E**, Beta values in **B** and cosine vector means in **E** of pattern similarities for item 1 and 2. Upper bars and corresponding shading indicate

significant values. Error shading indicates 95% C. I. of the mean. **C & F**, Topographies of each item of channel-wise decoding (100-400 ms relative to item onset).

Content-specific impulse responses

Auditory WM task

In the auditory impulse epoch, the neural dynamics time-course revealed significant cued-item decoding (180-308 ms, $p = 0.004$, one-sided, corrected), while no clusters were present for the uncued item (Fig. 3A & B, left). Similarly, the cued item was decodable in the visual impulse epoch (204-372 ms, $p = 0.009$, one-sided, corrected), while the uncued item was not (Fig. 3A & B, right).

The time-of-interest (100-400 ms relative to impulse onset) analysis provided similar results.

The cued item showed strong decoding in both impulse epochs (auditory impulse: Bayes factor = 11462.607, $p < 0.001$; visual impulse: Bayes factor = 85.843, $p < 0.001$, one-sided), but the uncued item did not (auditory impulse: Bayes factor = 0.968, $p = 0.075$; visual impulse: Bayes factor = 0.204, $p = 0.476$, one-sided; Fig. 3C). A model only including the cueing predictor yielded the highest Bayes factor of 8.123 (± 0.996 %) compared to the null model. A model including impulse modality as a predictor resulted in a Bayes factor of 0.848 (± 1.075 %). Including both predictors (impulse modality and cueing) in the model resulted in a Bayes factor of 7.553 (± 0.991 %) that was slightly lower than only including cueing.

Taken together, these results provided strong evidence that both impulse stimuli elicit neural responses that contain information about the cued item in auditory WM, but none about the uncued item.

Visual WM task

No significant time clusters were present in the auditory impulse epoch of the visual WM experiment for either the cued or the uncued item task (Fig. 3D & E, left). The decoding time-course of the visual impulse epoch revealed a significant decoding cluster of the cued item (108-396 ms, $p < 0.001$, one-sided, corrected) but not for the uncued item (Fig. 3D & E, right), replicating previous findings (Wolff et al., 2017).

The analysis on the time-of-interest interval (100-400 ms) showed the same pattern of results; neither the cued, nor uncued item in the auditory impulse epoch showed above 0 decoding (cued: Bayes factor = 0.236, $p = 0.417$; uncued: Bayes factor = 0.119, $p = 0.787$, one-sided).

In the visual impulse epoch the cued item showed strong decodability (Bayes factor = 1695.823, $p < 0.001$, one-sided) but the uncued item did not (Bayes factor = 0.236, $p = 0.421$, one-sided; Fig 3F). A model including both predictors (cueing and impulse modality) as well as their interaction resulted in the highest Bayes factor compared to the null model (Bayes factor = 56.284 ($\pm 1.557\%$)). Models with each predictor alone resulted in notably smaller Bayes factors (cueing: Bayes factor = 6.26 ($\pm 0.398\%$); impulse modality: Bayes factor = 5.877 ($\pm 0.686\%$)). The Bayes factor of the model including both predictors without interaction (46.728 ($\pm 0.886\%$)) was only 1.205 times smaller than the model that also included the interaction, highlighting that while there was strong evidence in favour of both impulse modality and cueing, there was only weak evidence in favour of an interaction.

Overall, these results provided evidence that while a visual impulse clearly evokes a neural response that contains information about the cued visual WM item, replicating previous findings (Wolff et al., 2017), an auditory impulse does not.

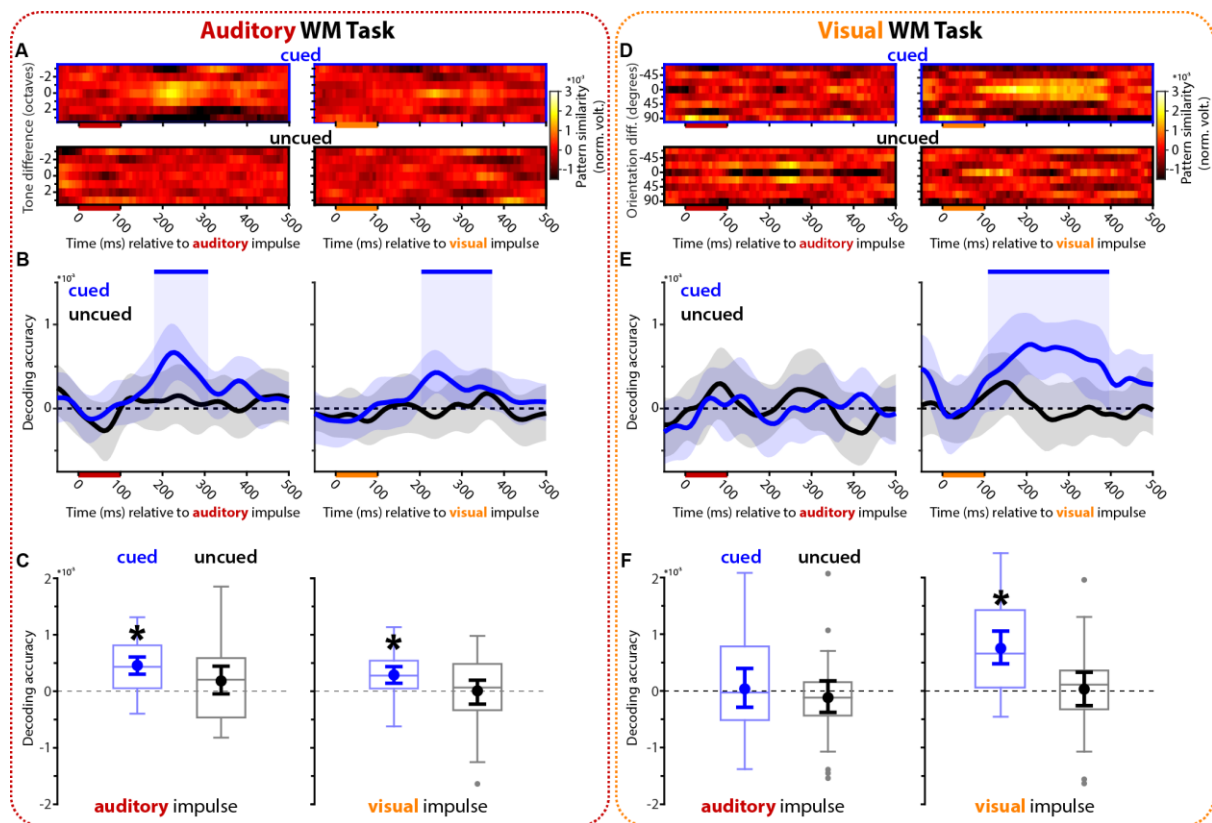


Figure 3. Decoding auditory and visual WM content from the impulse response

A-C, Auditory WM task. **D-F**, Visual WM task. **A & D**, Normalized average pattern similarity (mean-centred, sign-reversed mahalanobis distance) of the neural dynamics for each time-point between trials as a function of tone similarity in **A** and orientation similarity in **D**. Top row: cued item. Bottom row: uncued item. Left column: auditory impulse. Right column: visual impulse. **B & E**, Decoding accuracy time-course: Beta values in **B** and cosine vector means in **E** of pattern similarities for cued (blue) and uncued item (black). Upper bars and shading indicate significant values of the corresponding item. Error shading indicate 95% C. I. of the mean. **C & F**, Boxplots show the overall decoding accuracies for the cued (blue) and uncued (black) item, using the whole time-window of interest (100-400 ms relative to onset) from the auditory (left) and visual (right) impulse epoch. Centre lines indicate the median; box outlines show 25th and 75th percentiles, and whiskers indicate 1.5x the interquartile range. Extreme values are shown separately (dots). Superimposed circles and

error bars indicate mean and its 95% C.I., respectively. Asterisks indicate significant decoding accuracies ($p < 0.05$, one-sided).

Parametric encoding and maintenance of auditory pitch and visual orientation

As indicated, RSA was performed to explicitly test and explore for specific stimulus coding relationships in both experiments (Fig. 4A & D).

Auditory WM task

The RDMS of each epoch of interest are shown in Fig. 4B. There was strong evidence in favour of the pitch height difference model during item encoding (item 1 and item 2 presentation epochs; Bayes factor > 100.000 , $p < 0.001$, one-sided) while evidence against the pitch chroma model was evident (Bayes factor = 0.177, $p = 0.523$, one-sided; Fig. 4B & C, left). Moderate evidence in favour of the pitch height model was also evident for the cued item in the auditory impulse epoch (Bayes factor = 4.016, $p = 0.0113$, one sided), while there was weak evidence against the pitch chroma model (Bayes factor = 0.838, $p = 0.079$, one sided; Fig. 4B & C, middle). The visual impulse epoch also suggested a pitch height coding model of the cued auditory item, though the evidence was weak (Bayes factor = 1.346, $p = 0.049$, one sided), and there was again evidence against the pitch chroma model of the cued item (Bayes factor = 0.123, $p = 0.736$, one-sided; Fig. 4B & C, right).

Overall, these RSA results provide evidence that both the encoding and maintenance of pure tones are coded parametrically according to pitch height (Uluç et al., 2018), but not pitch chroma.

Visual WM task

474 The RDMs of the averaged encoding epochs (item 1 and item 2) and the visual impulse epoch
475 are shown in Fig. 4E. There was strong evidence in favour for a circular orientation
476 difference code (Bayes factor > 100.000 , $p < 0.001$, one-sided), as well as an additional
477 “cardinal specialization” code (Bayes factor > 100.000 , $p < 0.001$, one-sided) during item
478 encoding (Fig. 4E & F, left). The evoked neural response by the visual impulse also provided
479 strong evidence for a circular orientation difference code for the maintenance of the cued
480 item (Bayes factor = 362.672, $p < 0.001$, one-sided). No evidence in favour of an additional
481 “cardinal specialization” code during maintenance was found, however (Bayes factor =
482 0.252, $p = 0.318$, one-sided; Fig. 4E & F, right).

483 These results provide evidence that orientations are encoded and maintained in a parametric,
484 orientation selective code (e.g. Ringach, Shapley, & Hawken, 2002; Saproo & Serences,
485 2010). We additionally considered the “cardinal specialization” coding model, which
486 captures the expected increased neural distinctiveness of horizontal and vertical orientations
487 compared to tilted orientations, based on the superior visual discrimination of cardinal
488 orientations (Appelle, 1972) as well as previous neurophysiological reports of cardinal
489 specialization (Li, Peterson, & Freeman, 2003; Shen et al., 2014). Evidence for this model
490 was only found during orientation encoding, but not maintenance.

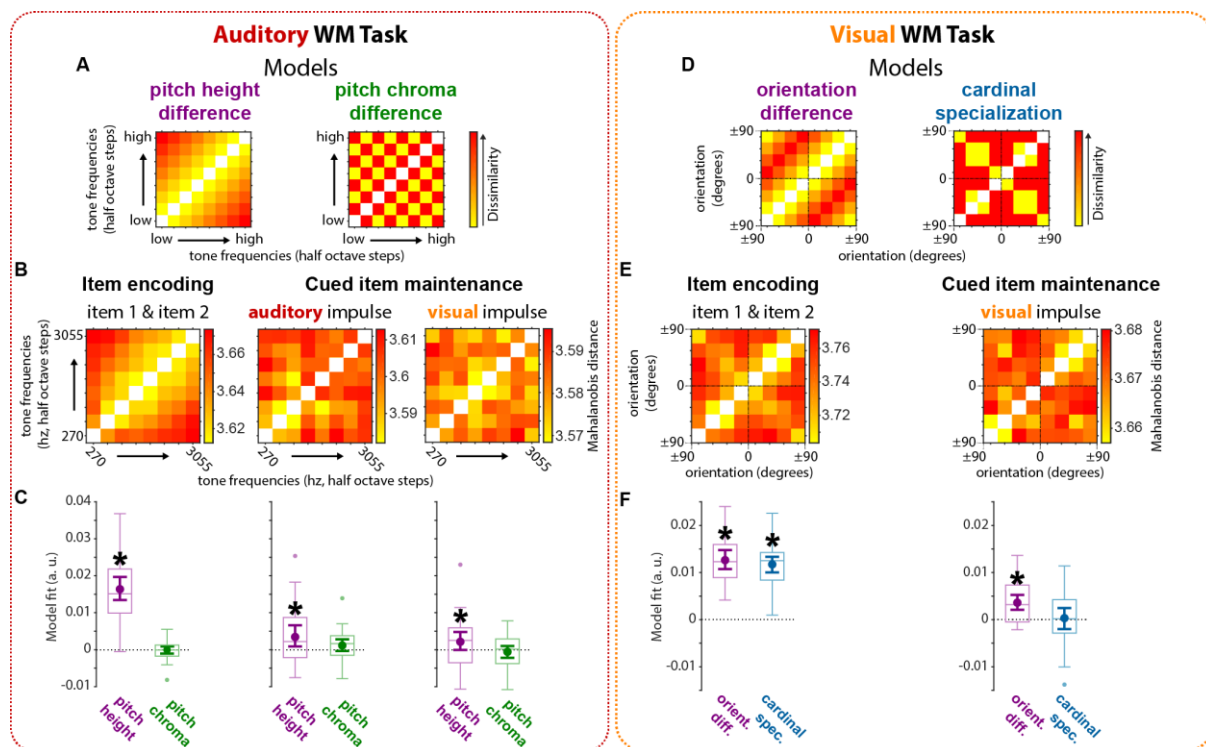


Figure 4. Stimulus coding relationship during encoding and maintenance

A-C, Auditory WM task. D-F, Visual WM task. A & D, Model RDMs of pitch (A) and orientation (D). B & E, Data RDMs. C & F, Model fits of model RDMs on data RDMs. Centre lines indicate the median; box outlines show 25th and 75th percentiles, and whiskers indicate 1.5x the interquartile range. Extreme values are shown separately (dots). Superimposed circles and error bars indicate mean its 95% C.I, respectively. Asterisks indicate significant model fits ($p < 0.05$, one-sided).

No WM-specific cross-generalization between impulse and WM-item presentation

It has been shown previously that the visual WM-dependent impulse response does not cross-generalize with visual item processing (Wolff et al., 2015). Here we tested if this is also the case for auditory WM, and additionally explored the cross-generalizability between impulses.

Auditory WM task

The representation of the cued item did neither cross-generalise between item presentation and either of the impulse epochs (auditory impulse: Bayes factor = 0.225, $p = 0.58$; visual impulse: Bayes factor = 0.356, $p = 0.26$, two-sided), nor between impulse epochs (Bayes factor = 0.267, $p = 0.417$, two-sided; Fig. 5A).

Visual WM task

Replicating previous reports (Wolff et al., 2015, 2017), the visual impulse response of the cued visual item did not cross-generalize with item processing during item presentation (Bayes factor = 0.491, $p = 0.168$, two-sided; Fig. 5B).

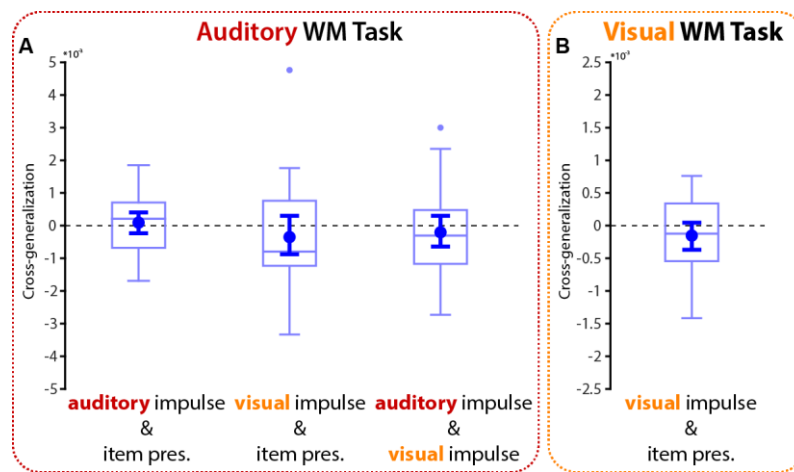


Figure 5. Cross-generalization between epochs

A, Cross-generalization of the cued item between the memory item epoch and impulse epochs in the auditory WM task. **B**, Cross-generalization between visual impulse and memory item in the visual WM task. Centre lines indicate the median; box outlines show 25th and 75th percentiles, and whiskers indicate 1.5x the interquartile range. Extreme values are shown separately (dots). Superimposed circles and error bars indicate mean and its 95% C.I. respectively.

522

523 **Evoked response magnitudes of impulse stimuli are comparable between tasks**

524 Since the impulse stimuli were always the same across trials, presented at the same relative
525 time within each trial, and were completely task irrelevant, we believe that the WM-specific
526 impulse responses reported here and in previous work rely on low-level interactions of the
527 impulse stimuli with the WM network, which do not depend on higher order cognitive
528 processing of the impulse.

529 Nevertheless, it could be argued that the impulse stimuli are differentially processed even at
530 an early stage between the WM tasks. Since the auditory impulse was the only auditory
531 stimulus in the visual WM task, it may have been more easily filtered out and ignored
532 compared to the other impulse stimuli. Indeed, it is possible that the neural response to the
533 auditory impulse stimulus was just too “weak” to result in a measurable, WM-specific neural
534 response in the visual WM task. However, given the uniqueness of the auditory impulse in
535 the visual WM task, the opposite could be argued as well.

536 To test for potential differences of attentional filtering of impulse stimuli between tasks, we
537 examined the event-related potentials (ERPs) to the impulse stimuli in both tasks from
538 electrodes associated with sensory processing (Fz, FCz, and Cz for auditory impulse; O1, Oz,
539 and O2 for visual impulse). If there is indeed a difference in early sensory processing, this
540 should be visible in associated early evoked responses within 250 ms of stimulus presentation
541 (Boutros, Korzyukov, Jansen, Feingold, & Bell, 2004; Luck, Woodman, & Vogel, 2000).
542 Because ERPs are subject to large individual differences, only participants who participated
543 in both tasks ($n = 16$) were included in this analysis.

544 We also considered potential voltage differences between tasks from 250 ms to 500 ms post
545 impulse onsets to test. This is the expected time range of the P3 ERP component and its two

subcomponents, the P3a and the P3b, which have been linked to the attentional processing of rare and unpredictable non-targets, and the processing (including memory consolidation) of target stimuli, respectively (Polich, 2007; Squires, Squires, & Hillyard, 1975). The presence of these components would imply that higher-order cognitive processes may be involved in the processing of the impulses, despite their regularity and task-irrelevance. To explore if the impulses elicited these endogenous components and test for potential differences between tasks, we considered the average voltages from channels Fz, FCz, and Cz for the P3a, and the average voltage from Pz for the P3b (Conroy & Polich, 2007).

Auditory ERPs

The early auditory ERP evoked from the auditory impulse stimulus within each task is shown in fig. 6A on the left. The P50, N1, and P2 components, all of which have been shown to be reduced when irrelevant auditory stimuli are filtered out (sensory gating; (e.g. Boutros et al., 2004; Cromwell, Mears, Wan, & Boutros, 2008; Kisley, Noecker, & Guinther, 2004), can clearly be identified in both tasks. One time-cluster of the difference between tasks was significant within the time-window of interest (148 to 184 ms, $p = 0.048$, two-sided, corrected). Visual inspection of the ERPs suggests that while there is no difference in P50 and N1 amplitude between tasks, P2 amplitude is larger in the visual than in the auditory task. Note that this difference goes in the opposite direction as would be expected if the auditory impulse stimulus was somehow more easily filtered out and ignored in the visual than in the auditory task.

The late ERP elicited by the auditory impulse stimuli in both tasks is shown in fig. 6A on the right. Visual inspection of the voltage traces suggests that no clear P3a or P3b components are evident, though it could be argued that the upward inflection at 300 ms in the frontal/central electrodes hints at a small P3a component (Fig. 6A, bottom right).

Nevertheless, no significant time-clusters in the difference between the auditory and the visual WM task were found in the time-window of interest in either voltage trace ($p > 0.19$, two-sided, corrected).

Visual ERPs

The early visual impulse ERP recorded from occipital electrodes is shown in fig. 6B on the left. Early components of interest (C1, P1, N1), which have been shown to be modulated by attentional processes (e.g. Di Russo, Martínez, & Hillyard, 2003; Luck et al., 2000; Rauss, Pourtois, Vuilleumier, & Schwartz, 2009), have been marked. Visual inspection suggests that there are no discernible differences in these visual components between tasks. Indeed, no significant time-clusters were found ($p > 0.19$, two-sided, corrected), suggesting that the visual impulse stimulus was processed similarly between tasks.

The late ERP in response to the visual impulse stimuli is shown in fig. 6B on the right. One significant time-cluster of the difference of the voltage traces between tasks was found in the frontal/central electrodes (266 to 322 ms, $p = 0.023$, two-sided, corrected; Fig. 6B bottom right). Visual inspection suggests that this could be due to a higher P3a amplitude in the visual than in the auditory task, implying that the visual impulse elicited more attentional processes. However, due to the generally small amplitude, a clear conclusion on what caused this difference cannot be drawn. Note that the visual impulse stimulus resulted in WM-specific responses in both tasks, so the observed voltage difference does not reconcile those findings. No time-clusters were found in the voltage difference between tasks on the posterior electrode (Fig. 6B, top right).

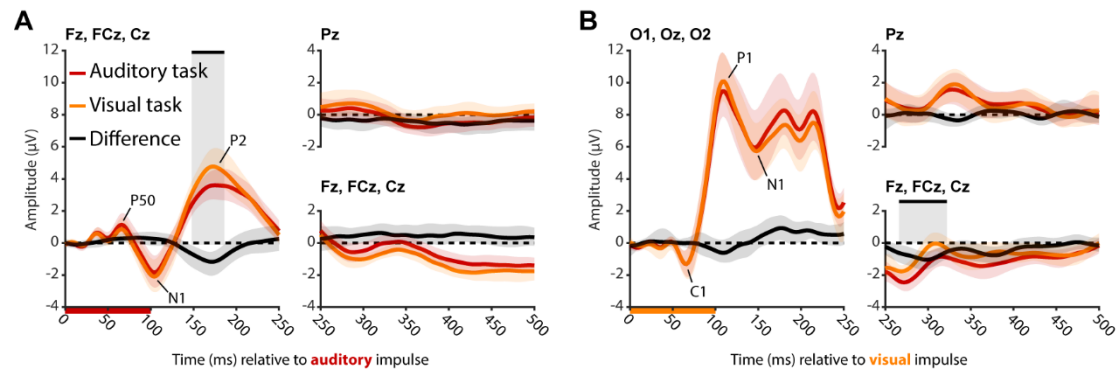


Figure 6. Evoked responses to impulse stimuli as a function of task for participants who participated in both tasks (n=16).

A, Average voltages evoked by auditory impulse in the auditory task (red) and visual task (orange). Difference voltage (auditory task minus visual task) is plotted in black. Individual ERP components of interest are labelled. Error shadings are 95% C. I. of the mean. The significant time-cluster of difference is indicated by the black bar ($p < 0.05$, corrected, two-sided). **B**, Average voltages evoked by the visual impulse. Same convention as in **A**.

Discussion

It has been shown that the bottom-up neural response to a visual impulse presented during the delay of a visual WM task contains information about relevant visual WM content (Wolff et al., 2015, 2017), which is consistent with WM theories that assume information is maintained in activity-silent brain states (Stokes, 2015). We used this approach to investigate whether sensory information is maintained within sensory-specific neural networks, shielded from other sensory processing areas. We show that the neural impulse response to sensory-specific stimulation is WM content-specific not only in visual WM, but also in auditory WM, demonstrating the feasibility and generalisability of the approach in the auditory domain. Furthermore, for auditory WM, a content-specific response was obtained not only during

auditory, but also during visual stimulation, suggesting a sensory modality-unspecific path to access the auditory WM network. In contrast, only visual, but not auditory, stimulation evoked a neural response containing relevant visual WM content. This pattern of impulse responsivity supports the idea that visual pathways may be more dominant in WM maintenance.

Recent studies have shown that delay activity in the auditory cortex reflects the content of auditory WM (Huang et al., 2016; Kumar et al., 2016; Uluç et al., 2018). Thus, similar to visual WM maintenance, which has been found to result in content-specific delay activity in the visual cortex (Harrison & Tong, 2009), auditory WM content is also maintained in a network that recruits the same brain area responsible for sensory processing. However, numerous visual WM studies have shown that content-specific delay activity may in fact reflect the focus of attention (Lewis-Peacock, Drysdale, Oberauer, & Postle, 2011; Sprague, Ester, & Serences, 2016; Watanabe & Funahashi, 2014). The memoranda themselves may instead be represented within connectivity patterns that generate a distinct neural response profile to internal or external neural stimulation (Lundqvist et al., 2016; Rose et al., 2016; Wolff et al., 2017). While previous research has focused on visual WM, we now provide evidence for a neural impulse response that reflects auditory WM content, suggesting a similar neural mechanism for auditory WM.

The neural response to a visual impulse stimulus also contained information about the behaviourally relevant pitch. It has been shown that visual stimulation can result in neural activity in the auditory cortex (Martuzzi et al., 2007; Morrill & Hasenstaub, 2018). Thus, direct connectivity between visual and auditory areas (Eckert et al., 2008) might be such that visual stimulation activates auditory WM representations in auditory cortex, providing an alternate access pathway. Alternatively, visual cortex itself might retain auditory information. It has been shown that natural sounds can be decoded from the activity in the visual cortex

during processing and imagination (Vetter, Smith, & Muckli, 2014). Even though pure tones were used in the present study, it is nevertheless possible that they have been visualised, for example by imagining the pitch as a location in space. Tones may have also resulted in semantic representations, by categorising them into arbitrary sets of low, medium, and high tones. The decodable signal from the impulse response might thus not necessarily originate from the sensory processing areas, but rather from higher brain regions such as the prefrontal cortex (Stokes et al., 2013). Future studies that employ imaging tools with high spatial resolution might be able to arbitrate the neural origin of the cross-modal impulse response in WM.

While the neural impulse response to visual stimulus contained information about the relevant visual WM item, replicating previous results (Wolff et al., 2017), the neural response to external auditory stimulation did not. This suggests that, in contrast to auditory information, visual information is maintained in a sensory-specific neural network with no evidence of content-specific connectivity with the auditory system, possibly reflecting the visual dominance of the human brain (Posner, Nissen, & Klein, 1976). Indeed, while it has been found that auditory stimulation results in neural activity in the visual cortex, it is notably weaker than the other way around (Martuzzi et al., 2007), which corresponds with our asymmetric findings of sensory specific and sensory non-specific impulse responses of visual and auditory WM.

One might argue that the asymmetric findings reported here could result from the asymmetry between experiments; whereas the auditory impulse was the only non-visual stimulus in the visual task, the auditory task contained several non-auditory stimuli (cue, fixation cross, visual impulse). The auditory impulse may have thus been more easily filtered out in the visual task, causing the neural response to be too “weak” to perturb the neural WM network. However, we found no evidence for this alternative explanation. None of the early sensory

auditory ERPs were smaller in amplitude in the visual task compared to the auditory task. Indeed, the auditory P2 was larger in the visual task, the opposite direction as would be expected if the auditory impulse was more easily ignored. There were furthermore no reliable differences in the early visual ERPs between tasks. In the later time window, there was no difference in the auditory ERPs either. The visual ERP at frontal electrodes did show elevated amplitude from 266 to 322 ms in the visual task, but the posterior electrode showed no difference. Perhaps most obvious was the lack of a clear P3 component in general, suggesting that the impulses did not elicit higher-level cognitive processing (see Polich, 2007 for review on P3). This is not unexpected, given their predictability and task-irrelevance in both tasks and modalities. Collectively, the ERPs do not support the idea that there might be systematic differences in impulse processing that could explain the differences in WM-specific impulse responses between tasks.

We found that both the processing and maintenance of pure tones was coded parametrically according to the height of the pitch, similar to previous reports of parametric auditory WM (Spitzer & Blankenburg, 2012; Uluç et al., 2018). On the other hand, a neural code for pitch chroma, the cyclical similarity of the same notes across different octaves, was not found during either perception or maintenance. It has previously been found that complex tones may be more likely to result in a neural representation of pitch chroma than pure tones (as were used in this study) during perception (Briley, Breakey, & Krumbholz, 2013).

Visual orientations were clearly coded parametrically during encoding and maintenance, replicating previous findings (e.g., Saproo & Serences, 2010). Interestingly, we also found evidence for a neural coding scheme that reflects the specialization of orientations close to the cardinal axes (horizontal and vertical) compared to the oblique orientations during the encoding of orientations. This coding scheme is related to the previously reported “oblique effect” (higher discrimination and report accuracy of cardinal compared to oblique

orientations; (Appelle, 1972)), and neural evidence for specialized neural structures in cat and macaque visual cortices for cardinal orientations (Li et al., 2003; Shen et al., 2014). The visual impulse response did not reveal such a coding scheme during maintenance, however, which could reflect a genuinely different coding scheme, but could also be due to the generally weaker orientation code during maintenance.

It has been reported that the WM-related neural pattern evoked by the impulse response does not cross-generalize with the neural activity evoked by the memory stimulus itself (Wolff et al., 2015), suggesting that the neural activation patterns are qualitatively different. In the present study, we also found no cross-generalization between item processing and the impulse response, neither in the visual nor in the auditory WM task. The neural representation of WM content may thus not be an exact copy of stimulation history, literally reflecting the activity pattern during information processing and encoding, but rather a reconfigured code that is optimized for future behavioural demands (Myers, Stokes, & Nobre, 2017). Similarly, no generalizability was found between auditory and visual impulse responses in the auditory task. This could suggest that distinct neural networks are perturbed by the different impulse modalities, or, as alluded to above, that it reflects the unique interaction between impulses and the perturbed neural network. Future research should employ neural imaging tools with high spatial resolution to investigate the neural populations involved in the WM-dependent impulse response.

The present results provide a novel approach to the ongoing debate on the extent to which sensory processing areas are essential for the maintenance of information in WM (Gayet, Paffen, & Van der Stigchel, 2018; Scimeca et al., 2018; Xu, 2018). This is usually investigated by measuring WM-specific delay activity in the visual cortex in visual WM tasks (Bettencourt & Xu, 2016; Harrison & Tong, 2009), where null results are interpreted as evidence against the involvement of specific brain regions, which is inherently problematic

711 (Ester, Rademaker, & Sprague, 2016), and by which non-active WM states are not
712 considered. In the present study, we found that sensory-specific stimulation, and both sensory
713 specific and non-specific stimulation, resulted in WM-specific neural responses during the
714 maintenance of visual and auditory information, respectively. Sensory cortices were thus
715 linked to WM maintenance not by relying on ambient delay activity, but rather by perturbing
716 the underlying, connectivity-dependent, representational WM network via a bottom-up neural
717 response.

References

- Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The “oblique effect” in man and animals. *Psychological Bulletin*, 78(4), 266–278.
- Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, 19(1), 150–157. <https://doi.org/10.1038/nn.4174>
- Boutros, N. N., Korzyukov, O., Jansen, B., Feingold, A., & Bell, M. (2004). Sensory gating deficits during the mid-latency phase of information processing in medicated schizophrenia patients. *Psychiatry Research*, 126(3), 203–215. <https://doi.org/10.1016/j.psychres.2004.01.007>
- Briley, P. M., Breakey, C., & Krumbholz, K. (2013). Evidence for Pitch Chroma Mapping in Human Auditory Cortex. *Cerebral Cortex*, 23(11), 2601–2610. <https://doi.org/10.1093/cercor/bhs242>
- Buonomano, D. V., & Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2), 113–125. <https://doi.org/10.1038/nrn2558>
- Chang, A., Bosnyak, D. J., & Trainor, L. J. (2016). Unpredicted Pitch Modulates Beta Oscillatory Power during Rhythmic Entrainment to a Tone Sequence. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00327>
- Conroy, M. A., & Polich, J. (2007). Normative Variation of P3a and P3b from a Large Sample: Gender, Topography, and Response Time. *Journal of Psychophysiology* 2007, 21(1), 22–32. <https://doi.org/10.1027/0269-8803.21.1.22>
- Cromwell, H. C., Mears, R. P., Wan, L., & Boutros, N. N. (2008). Sensory Gating: A Translational Effort from Basic to Clinical Science. *Clinical EEG and Neuroscience*, 39(2), 69–72. <https://doi.org/10.1177/155005940803900209>

743 De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis
744 distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1), 1–18.
745 [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7)

746 Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-
747 trial EEG dynamics including independent component analysis. *Journal of*
748 *Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>

749 Derrick, B., Toher, D., & White, P. (2017). How to compare the means of two samples that
750 include paired observations and independent observations: A companion to Derrick,
751 Russ, Toher and White (2017). *The Quantitative Methods for Psychology*, 13(2), 120–
752 126. <https://doi.org/10.20982/tqmp.13.2.p120>

753 Di Russo, F., Martínez, A., & Hillyard, S. A. (2003). Source Analysis of Event-related
754 Cortical Activity during Visuo-spatial Attention. *Cerebral Cortex*, 13(5), 486–499.
755 <https://doi.org/10.1093/cercor/13.5.486>

756 Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends*
757 *in Cognitive Sciences*, 2(7), 254–262. [https://doi.org/10.1016/S1364-6613\(98\)01188-7](https://doi.org/10.1016/S1364-6613(98)01188-7)

758 Eckert, M. A., Kamdar, N. V., Chang, C. E., Beckmann, C. F., Greicius, M. D., & Menon, V.
759 (2008). A Cross-Modal System Linking Primary Auditory and Visual Cortices.
760 *Human Brain Mapping*, 29(7), 848–857. <https://doi.org/10.1002/hbm.20560>

761 Ester, E. F., Rademaker, R. L., & Sprague, T. C. (2016). How Do Visual and Parietal Cortex
762 Contribute to Visual Short-Term Memory? *ENeuro*, 3(2), ENEURO.0041-16.2016.
763 <https://doi.org/10.1523/ENeuro.0041-16.2016>

764 Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2018). Visual Working Memory Storage
765 Recruits Sensory Processing Areas. *Trends in Cognitive Sciences*, 22(3), 189–190.
766 <https://doi.org/10.1016/j.tics.2017.09.011>

767 Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns
768 from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time
769 Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, 29(4), 677–697.
770 https://doi.org/10.1162/jocn_a_01068

771 Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory
772 in early visual areas. *Nature*, 458(7238), 632–635.
773 <https://doi.org/10.1038/nature07832>

774 Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G., & Nelson, S. B. (2000).
775 Multiple Forms of Short-Term Plasticity at Excitatory Synapses in Rat Medial
776 Prefrontal Cortex. *Journal of Neurophysiology*, 83(5), 3031–3041.

777 Huang, Y., Matysiak, A., Heil, P., König, R., & Brosch, M. (2016). Persistent neural activity
778 in auditory cortex is related to auditory working memory in humans and nonhuman
779 primates. *ELife*, 5. <https://doi.org/10.7554/eLife.15441>

780 Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., ... Medini, P. (2012).
781 Sound-Driven Synaptic Inhibition in Primary Visual Cortex. *Neuron*, 73(4), 814–828.
782 <https://doi.org/10.1016/j.neuron.2011.12.026>

783 Kisley, M. A., Noecker, T. L., & Guinther, P. M. (2004). Comparison of sensory gating to
784 mismatch negativity and self-reported perceptual phenomena in healthy adults.
785 *Psychophysiology*, 41(4), 604–612. <https://doi.org/10.1111/j.1469-8986.2004.00191.x>

786 Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—
787 Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*,
788 2. <https://doi.org/10.3389/neuro.06.004.2008>

789 Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., & Griffiths, T. D. (2016).
790 A Brain System for Auditory Working Memory. *Journal of Neuroscience*, 36(16),
791 4492–4505. <https://doi.org/10.1523/JNEUROSCI.4341-14.2016>

792 Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of*
 793 *Portfolio Management*, 30(4), 110–119. <https://doi.org/10.3905/jpm.2004.110>
 794 Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2011). Neural Evidence
 795 for a Distinction between Short-term Memory and the Focus of Attention. *Journal of*
 796 *Cognitive Neuroscience*, 24(1), 61–79. https://doi.org/10.1162/jocn_a_00140
 797 Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique Effect: A Neural Basis in the
 798 Visual Cortex. *Journal of Neurophysiology*, 90(1), 204–217.
 799 <https://doi.org/10.1152/jn.00954.2002>
 800 Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of
 801 attention. *Trends in Cognitive Sciences*, 4(11), 432–440.
 802 [https://doi.org/10.1016/S1364-6613\(00\)01545-X](https://doi.org/10.1016/S1364-6613(00)01545-X)
 803 Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016).
 804 Gamma and Beta Bursts Underlie Working Memory. *Neuron*, 90(1), 152–164.
 805 <https://doi.org/10.1016/j.neuron.2016.02.028>
 806 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.
 807 *Journal of Neuroscience Methods*, 164(1), 177–190.
 808 <https://doi.org/10.1016/j.jneumeth.2007.03.024>
 809 Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J.-P., Maeder, P. P., Clarke, S., &
 810 Meuli, R. A. (2007). Multisensory Interactions within Human Primary Cortices
 811 Revealed by BOLD Dynamics. *Cerebral Cortex*, 17(7), 1672–1679.
 812 <https://doi.org/10.1093/cercor/bhl077>
 813 Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic Theory of Working Memory.
 814 *Science*, 319(5869), 1543–1546. <https://doi.org/10.1126/science.1150769>

815 Morrill, R. J., & Hasenstaub, A. R. (2018). Visual Information Present in Infragranular
816 Layers of Mouse Auditory Cortex. *Journal of Neuroscience*, 38(11), 2854–2862.
817 <https://doi.org/10.1523/JNEUROSCI.3102-17.2018>

818 Myers, N. E., Rohenkohl, G., Wyart, V., Woolrich, M. W., Nobre, A. C., & Stokes, M. G.
819 (2015). Testing sensory evidence against mnemonic templates. *ELife*, 4, e09000.
820 <https://doi.org/10.7554/eLife.09000>

821 Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working
822 Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences*.
823 <https://doi.org/10.1016/j.tics.2017.03.010>

824 Nemrodov, D., Niemeier, M., Patel, A., & Nestor, A. (2018). The Neural Dynamics of Facial
825 Identity Processing: Insights from EEG-Based Pattern Analysis and Image
826 Reconstruction. *ENeuro*, 5(1), ENEURO.0358-17.2018.
827 <https://doi.org/10.1523/ENeuro.0358-17.2018>

828 Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., Oostenveld, R., Fries, P., ...
829 Schoffelen, J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of
830 MEG, EEG, and Invasive Electrophysiological Data, FieldTrip: Open Source
831 Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological
832 Data. *Computational Intelligence and Neuroscience, Computational Intelligence and*
833 *Neuroscience*, 2011, 2011, e156869. <https://doi.org/10.1155/2011/156869>,
834 10.1155/2011/156869

835 Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical*
836 *Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>

837 Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-
838 processing account of its origins and significance. *Psychological Review*, 83(2), 157–
839 171.

- Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 6–17. <https://doi.org/10.1037/xhp0000302>
- Rauss, K. S., Pourtois, G., Vuilleumier, P., & Schwartz, S. (2009). Attentional load modifies early activity in human primary visual cortex. *Human Brain Mapping*, 30(5), 1723–1733. <https://doi.org/10.1002/hbm.20636>
- Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. *Journal of Neuroscience*, 22(13), 5639–5651. <https://doi.org/10.1523/JNEUROSCI.22-13-05639.2002>
- 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. <https://doi.org/10.1126/science.aah7011>
- Saproo, S., & Serences, J. T. (2010). Spatial Attention Improves the Quality of Population Codes in Human Visual Cortex. *Journal of Neurophysiology*, 104(2), 885–895. <https://doi.org/10.1152/jn.00369.2010>
- Scimeca, J. M., Kiyonaga, A., & D’Esposito, M. (2018). Reaffirming the Sensory Recruitment Account of Working Memory. *Trends in Cognitive Sciences*, 22(3), 190–192. <https://doi.org/10.1016/j.tics.2017.12.007>
- Shen, G., Tao, X., Zhang, B., Smith, E. L., & Chino, Y. M. (2014). Oblique effect in visual area 2 of macaque monkeys. *Journal of Vision*, 14(2), 3–3. <https://doi.org/10.1167/14.2.3>

863 Spitzer, B., & Blankenburg, F. (2012). Supramodal Parametric Working Memory Processing
864 in Humans. *Journal of Neuroscience*, 32(10), 3287–3295.
865 <https://doi.org/10.1523/JNEUROSCI.5280-11.2012>

866 Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring Latent Visual Working
867 Memory Representations in Human Cortex. *Neuron*, 91(3), 694–707.
868 <https://doi.org/10.1016/j.neuron.2016.07.006>

869 Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency
870 positive waves evoked by unpredictable auditory stimuli in man.
871 *Electroencephalography and Clinical Neurophysiology*, 38(4), 387–401.
872 [https://doi.org/10.1016/0013-4694\(75\)90263-1](https://doi.org/10.1016/0013-4694(75)90263-1)

873 Stokes, M. G. (2015). ‘Activity-silent’ working memory in prefrontal cortex: A dynamic
874 coding framework. *Trends in Cognitive Sciences*, 19(7), 394–405.
875 <https://doi.org/10.1016/j.tics.2015.05.004>

876 Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic
877 Coding for Cognitive Control in Prefrontal Cortex. *Neuron*, 78(2), 364–375.
878 <https://doi.org/10.1016/j.neuron.2013.01.039>

879 Uluç, I., Schmidt, T. T., Wu, Y., & Blankenburg, F. (2018). Content-specific codes of
880 parametric auditory working memory in humans. *NeuroImage*, 183, 254–262.
881 <https://doi.org/10.1016/j.neuroimage.2018.08.024>

882 Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding Sound and Imagery Content in Early
883 Visual Cortex. *Current Biology*, 24(11), 1256–1262.
884 <https://doi.org/10.1016/j.cub.2014.04.020>

885 Watanabe, K., & Funahashi, S. (2014). Neural mechanisms of dual-task interference and
886 cognitive capacity limitation in the prefrontal cortex. *Nature Neuroscience*, 17(4),
887 601–611. <https://doi.org/10.1038/nn.3667>

888 Wessel, J. R., & Aron, A. R. (2017). On the Globality of Motor Suppression: Unexpected
889 Events and Their Influence on Behavior and Cognition. *Neuron*, 93(2), 259–280.
890 <https://doi.org/10.1016/j.neuron.2016.12.013>

891 Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in
892 visual working memory using electroencephalography. *Frontiers in Systems*
893 *Neuroscience*, 9. <https://doi.org/10.3389/fnsys.2015.00123>

894 Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states
895 underlying working-memory-guided behavior. *Nature Neuroscience*, 20(6), 864–871.
896 <https://doi.org/10.1038/nn.4546>

897 Xu, Y. (2017). Reevaluating the Sensory Account of Visual Working Memory Storage.
898 *Trends in Cognitive Sciences*, 21(10), 794–815.
899 <https://doi.org/10.1016/j.tics.2017.06.013>

900 Xu, Y. (2018). Sensory Cortex Is Nonessential in Working Memory Storage. *Trends in*
901 *Cognitive Sciences*, 22(3), 192–193. <https://doi.org/10.1016/j.tics.2017.12.008>