Patterns of information segregation during working memory and attention revealed by dual-task interference in behavior, pupillometry, and EEG

Abbreviated Title: Information segregation in working memory and attention

Justin T. Fleming[1], J. Michelle Njoroge[2], Abigail L. Noyce[3], Tyler K. Perrahione[1,2], and Barbara G. Shinn-Cunningham[3]

1. Program in Speech and Hearing Bioscience and Technology, Harvard University, Cambridge, MA 02138
2. Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215
3. Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA 15213

Corresponding author: Justin T. Fleming (jtf@umn.edu)

Number of Figures: 9

The authors declare no competing interests.

Project Abstract

Making sense of our environment requires us to extract simultaneous temporal and spatial information from multiple sensory modalities, particularly audition and vision. This sensory information can be stored in working memory (WM) to guide future actions, at which point it must be safeguarded against interference from ongoing sensory processing. Recent fMRI research has uncovered regions in human frontal cortex well-suited to coordinate this interplay between attention and WM for multisensory and multidimensional information. Which of these brain regions are engaged depends on both the sensory modality of the input and the information domain of the task, forming the basis of two complementary networks specialized for auditory/temporal and visual/spatial processing. Motivated by the functional specializations of these networks, we examined whether similarity in sensory modality and information domain modulates neural and perceptual interference between two concurrent tasks. Participants stored temporal or spatial information about auditory or visual stimuli in WM, and on some trials, performed an intervening temporal or spatial auditory task during WM retention. WM recall and auditory perceptual judgments were impaired when the two tasks relied on the same sensory modality and/or information domain. Pupil dilations were also larger in these conditions, indicating increased cognitive effort. Event-related potentials (ERPs) revealed a neural signature of domain-based interference that was masked by behavioral ceiling effects. These results demonstrate that modality and information domain jointly affect how task information is represented in WM, and concomitantly, how tasks engage the complementary auditory-temporal and visual/spatial cognitive control networks.

Keywords: Spatial, temporal, audio-visual, dual-task interference, working memory, attention, pupillometry, EEG, alpha power, lateral frontal cortex (LFC), multiple resource theory

Introduction

Navigating everyday environments requires us to process spatial and temporal information from multiple sensory modalities, particularly vision and audition. Once stored in working memory (WM), these perceptual inputs can guide future actions. For instance, when crossing a busy street, we first look in one direction, sampling the spatial positions of vehicles over time to determine their direction and speed of movement. This information then moves into WM while we look the other way to determine the paths of vehicles approaching from that direction. To determine when it is safe to cross, we must maintain the temporal and spatial information stored in WM while simultaneously processing additional sensory inputs.

Under ideal circumstances, ongoing perceptual processing would not disrupt information stored in WM, and vice versa. Unfortunately, interference between WM maintenance and a concurrent task occurs all the time, especially when the information in WM and the sensory inputs from a secondary task share common attributes. For example, participants experience increased word-color Stroop interference when WM is loaded with similar verbal information, but not when WM is loaded with unrelated spatial information (Kim et al., 2005). Other studies featuring a WM task and distractor stimuli have shown that distractors selectively interfere with recall of closely related items from WM. For instance, in visual-spatial WM, distractors that vary in polar angle selectively impair recall of a target polar angle (Marini et al., 2017); in verbal WM, distractor words selectively interfere with participants' ability to recall phonologically similar target words (Oberauer & Lange, 2008). These findings support the "multiple resource theory" account of dual-task interference, which posits that two tasks are more likely to interfere with each other when they draw on a shared pool of specific cognitive resources (Navon & Gopher, 1979; Nickerson, 1980; Wickens, 2002).

Concurrent tasks can also interfere with one another if they require processing information from the same sensory modality. In dual-task paradigms, increased interference has been observed when both tasks are auditory, visual, or tactile compared to conditions in which the tasks operate on inputs from different sensory modalities (Morrison et al., 2015; Scerra & Brill, 2012). Additionally, auditory and visual discrimination thresholds (for pitch and contrast, respectively) are unaffected by concurrent, irrelevant stimuli in the opposite modality, but are considerably worse in the presence of concurrent stimuli in the same modality (Alais et al., 2006).

Recent fMRI research has identified subregions within human frontal cortex (LFC) that are preferentially driven by auditory or visual information during attention and WM tasks (Braga et al., 2017; Mayer et al., 2016; Michalka et al., 2015; Noyce et al., 2017). These LFC subregions are strongly connected to the corresponding primary sensory brain areas, forming networks tuned to processing either auditory or visual information and storing it in WM (Michalka et al., 2015; Tobyne et al., 2017). The organization of these networks may explain the patterns of interference based on sensory modality reported in previous studies; an auditory and a visual task should drive separate sensory-biased networks and therefore interfere relatively little with one another, whereas tasks presented in the same sensory modality should interfere more.

Although each LFC network is biased toward information from one sensory modality, each network can also be recruited to process information from the non-preferred sensory modality. This occurs in a manner consistent with the complementary specializations of vision for spatial processing and audition for temporal processing. From the retina, visual representations are inherently spatial, and neural maps of space are found throughout the visual processing pathway (Silver & Kastner, 2009; Stensaas et al., 1974; Swisher et al., 2007). In contrast, auditory spatial information must be indirectly computed by comparing the signals reaching the two ears. The

92 peripheral auditory system instead excels at representing temporal information; each auditory
93 nerve fiber, tuned to a narrow acoustic frequency range, phase-locks to the temporal structure of
94 the portion of the sound pressure wave to which it responds (Dynes & Delgutte, 1992). Reflecting
95 these complementary strengths, auditory tasks that involve *spatial* information processing recruit
96 the visual-biased LFC network, while visual tasks that involve *temporal* judgments recruit the
97 auditory-biased LFC network (Michalka et al., 2015). This flexible allocation of resources allows
98 information from either sensory modality to be processed by a network specialized for the temporal
99 and spatial strengths of audition and vision, respectively.

100 The functional specializations of these networks point to a potential neural substrate for the
101 "resources" in multiple resource theory. In a dual-task setting (such as performing a perceptual
102 task under WM load), two tasks that rely on the same LFC network should interfere strongly with
103 each other, whereas tasks that rely on different networks should interfere less. In addition to
104 explaining the greater interference between tasks in the same sensory modality, this pattern of
105 cross-modal LFC network recruitment yields a previously untested prediction: that dual-task
106 interference may depend on interactions between sensory modality *and* information domain (e.g.,
107 whether the task is temporal vs. spatial). For example, performing an auditory-temporal perceptual
108 task should strongly interfere with auditory-temporal information stored in WM. On the other
109 hand, the same perceptual task might interfere less with auditory-*spatial* WM; even though both
110 tasks operate on auditory inputs, auditory-spatial information may be stored and maintained in the
111 visual-biased network, resulting in less interference with auditory-temporal perceptual processing.
112 The present work tests the idea that, in addition to similarity in sensory modality, interference
113 between WM and perceptual tasks will be modulated by similarity in information domain. We
114 developed a dual-task interference paradigm featuring WM and "Intervening" tasks. The WM task

required participants to remember either temporal or spatial information about a set of nonverbal auditory or visual stimuli. The Intervening task, which was presented during WM retention, required participants to make an immediate perceptual judgment about either the timing or spatial locations of auditory stimuli. This paradigm allowed us to observe patterns of interference as participants engaged a WM network to encode sensory information in memory, performed a real-time perceptual task that required either increasing the load on the active network or switching networks, and finally retrieved the information stored in WM.

In addition to behavioral performance on the Intervening and WM tasks, we simultaneously measured pupillometry and electroencephalography (EEG) as participants encoded and held information in WM. Pupillometry serves as a sensitive psychophysiological index of the cognitive effort required to achieve a certain performance level, even in the absence of behavioral differences. For instance, maximum pupil dilations evoked by speech are larger for listeners with hearing loss than normal-hearing controls, even at high signal-to-noise ratios where both groups perform at ceiling (Ohlenforst et al., 2017). Under some circumstances, pupil dilations also correlate with participants' subjective assessments of task difficulty (Koelewijn et al., 2015). As a time-series measure, pupillometry can also provide insights into *when* participants deployed effort over the course of each dual-task trial. The EEG data provided an even more temporally resolved signal, allowing us to examine the effects of dual-task interference on the neural responses to individual stimulus events (event-related potentials, ERPs) and an established oscillatory signature of WM maintenance (alpha power in the 8-13 Hz frequency band). Across these measures, we expected to see evidence of heightened interference when the Intervening and WM tasks were matched in sensory modality and/or information domain. We hypothesized that interference in these conditions would manifest as increased behavioral errors and larger task-evoked pupil

138 dilations. Among the multiple possible effects that dual-task interference might have on EEG
139 signals (Lin et al., 2011; Obleser et al., 2012; Serrien et al., 2004), we were particularly interested
140 in testing the following predictions: As interference increased on the basis of sensory modality and
141 information domain, ERP amplitudes would be reduced (reflecting taxation of resources shared
142 between the two tasks) and ongoing alpha oscillations during WM maintenance would be
143 interrupted.
144

145 **Materials and Methods**

146 *Participants*

147 Twenty-three healthy young adults completed all experimental procedures. Data from three
148 participants were removed due to excessive noise in the pupillometry or EEG recordings, making
149 for a final sample of *N*=20 (13 female; mean age 20.9 years; range 18 to 28 years). This recruitment
150 target was based on a power analysis conducted on behavioral pilot data (N=6), which indicated
151 that 15 participants would be required to detect differences in auditory perceptual performance
152 based on the type of information held in WM. To account for variability between participants and
153 the potential for additional noise in the EEG and pupillometry data, the recruitment goal was
154 increased to 20.

155 One participant was excluded from only the time-frequency analyses due to anomalous
156 high-frequency noise in their EEG data. All participants had normal or corrected-to-normal visual
157 acuity and no reported colorblindness. Participants with corrected vision wore contact lenses
158 instead of glasses to avoid potential artifacts in the pupillometry data. All participants had
159 clinically normal hearing, defined by tone detection thresholds below 20 dB HL at octave
160 frequencies between 250 Hz and 8 kHz, as confirmed by an audiometric screening. Participants

gave written informed consent and were compensated for their participation. All study procedures were approved and overseen by the Boston University Charles River Campus Institutional Review Board.

Experimental Setup

The experiment was conducted in a darkened, electrically shielded, sound-treated booth. Participants were seated comfortably with their chin resting on a desk-mounted head support (SR Research). A BenQ 1080p LED monitor (27-inch diagonal, 120 Hz refresh rate) was positioned in front of the participant at approximately 65 cm distance. The monitor was set to 3% of its maximum brightness level to prevent eye fatigue and pupil diameter saturation. An EyeLink 1000 Plus eye-tracking system was placed on the desk just below the display for measurement of pupil diameter. Six free-field loudspeakers (KEF E301) were mounted in an arc around the participant at a distance of 1.5 m. Five of the loudspeakers were equally spaced in azimuth at ±90°, ±45°, and 0° relative to midline; the sixth was placed immediately to the left of the central loudspeaker, at −4° azimuth, and used only in the Intervening tasks described below (Fig. 1A). All six loudspeakers were positioned at approximately 5° elevation relative to the horizontal plane of the eyes to reduce obstruction by the visual display. Auditory stimulus presentation was handled using an RME Fireface UCX soundcard.

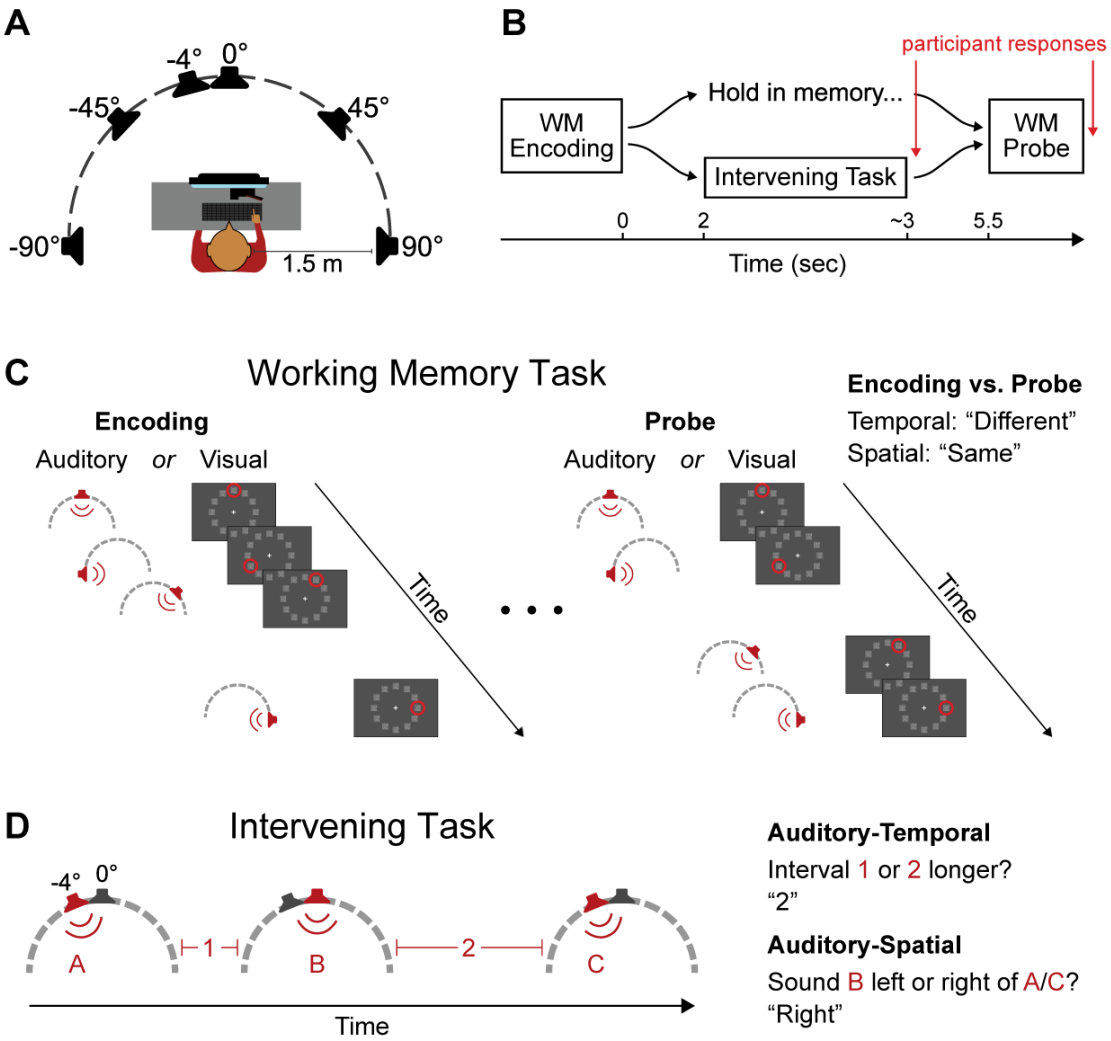

Figure 1: Experimental setup and task design. A, Overhead depiction of the experimental setup. All the loudspeakers except the one at −4° azimuth were used for the WM tasks, while only the loudspeakers at −4° and 0° were used for the Intervening tasks. **B**, Overall dual-task structure. The time base is relative to the offset of the final stimulus in the WM encoding phase. **C**, WM task structure. Loudspeakers playing each auditory stimulus and the changing visual stimuli are shown in red. Sensory modality always matched between the encoding and probe sequences. Correct responses for this example trial, shown on the right, differ depending on whether task was temporal or spatial. Note that in the actual experiment, inter-stimulus intervals were always isochronous in the auditory-spatial WM task. **D**, Intervening task structure. Auditory stimuli and inter-stimulus intervals are numbered and referred to in the temporal and spatial task responses on the right.

A standard keyboard was used to register all task responses. 64-channel EEG data were collected at a sampling rate of 2048 Hz using a Biosemi ActiveTwo system. Separate PCs were used for pupillometry recording and EEG recording, and a third PC was used for presenting stimuli

and registering behavioral responses. To ensure synchrony of event triggers (e.g., trial starts, stimulus presentation) between the pupillometry and EEG data, triggers were output through the S/PDIF channel on the soundcard, converted to TTL pulses using a custom converter box, and written simultaneously into the EEG and pupillometry data files. Experiment control was carried out using custom MATLAB software, and visual stimulus presentation was implemented using the Psychtoolbox package (Brainard, 1997).

Task and Experimental Design

Participants performed a dual-task paradigm, comprising a working memory (WM) task and an Intervening task (Fig. 1B). Each trial started with a 1.5-s baseline period, followed by the presentation of a sequence of four auditory or visual stimuli to be encoded in WM. Each stimulus was presented at one of five (auditory) or twelve (visual) locations, and each inter-stimulus interval in the sequence was randomly set to be either short or long (more details below). The WM task could be either temporal or spatial, yielding four total WM task conditions: auditory-temporal (AT), auditory-spatial (AS), visual-temporal (VT), and visual-spatial (VS). When the WM task was temporal, participants were instructed to remember the pattern of inter-stimulus intervals (i.e., the rhythm), regardless of spatial locations. When the WM task domain was spatial, participants had to remember the locations of the stimuli, regardless of order or timing (Fig. 1C). In all but the AS WM task (see below), both the locations and intervals could change between the encoding and probe sequences, but participants were instructed to ignore changes in the unattended domain. These WM task conditions were derived from tasks used to recruit the complementary LFC networks in fMRI studies (Michalka et al., 2015).

Participants retained stimulus information in WM for 5.5 seconds, after which a four-stimulus probe sequence was presented in the same sensory modality as the encoded sequence. Participants compared the encoded and probe sequences and made a same-different judgment on the remembered domain (temporal or spatial). After the conclusion of the probe stimulus, participants had 1.5 sec to indicate whether the encoded and probe sequences were the same (by pressing "1" on the keyboard) or different (by pressing "0"). Each block contained an equal number of same and different trials, ordered randomly. Participants maintained fixation on a small black cross (0.41° visual angle) at the center of the display throughout the trial.

We conducted pilot testing of the WM tasks to determine parameter settings that avoided ceiling or floor effects in behavioral measures. For participants to perceive the different inter-stimulus intervals equally well, a larger temporal separation was needed for the visual stimuli (200 and 580 ms) than the auditory stimuli (200 and 340 ms). Conversely, the visual-spatial task was too easy with only five stimulus locations, so the number of potential visual locations was increased to 12. Finally, participants struggled to perform the AS WM task when stimulus timing was variable; therefore, both encoding and probe stimuli in this condition were presented isochronously at the longer inter-stimulus interval.

On some trials, participants also performed an Intervening task during the WM retention period. This task was always auditory to allow pupil diameter to be measured in the absence of any visual stimulation, but like the WM tasks, it could be either temporal or spatial (AT or AS; see Fig. 1D). The stimulus structure was the same for the temporal and spatial variants. Starting 2 sec after the offset of the final stimulus in the WM encoding phase, a sequence of three auditory stimuli was presented. These stimuli were white noise bursts, acoustically distinct from the stimuli used in the auditory WM tasks (tone complexes) to prevent confusion between the tasks. One of the two

intervals between the stimuli was randomly chosen to be slightly longer than the other. The precise intervals were jittered on each trial, with an average interval duration of 460 ms and a 90 ms average difference between the two intervals. The sounds were presented from the two near-frontal loudspeakers (−4° and 0° azimuth). The first stimulus played from one of these loudspeakers, chosen randomly and with equal probability; the second stimulus was played from the other loudspeaker, and the third was played from the same location as the first. In the temporal Intervening task, participants judged whether the first or second inter-stimulus interval was longer. In the spatial Intervening task, participants were asked to determine whether the second sound was to the left or right relative to the first and third sounds.

This stimulus design allowed physically identical auditory stimuli to be used for the spatial and temporal Intervening task conditions. However, it did introduce an asymmetry between conditions in the amount of information required to do the task. In the temporal Intervening task, participants needed to attend all three stimuli in order to compare the two inter-stimulus intervals, whereas in the spatial Intervening task, participants were often able to make their judgment on the second auditory stimulus by comparing its location to the first. This could result in a longer period of increasing pupil size in the temporal task, leading to larger peak pupil diameter. In addition, behavioral accuracy differed between the two Intervening tasks. Both of these differences could confound comparisons across the different Intervening tasks; however, our analyses focused mainly on comparisons across the four WM conditions *within* each Intervening task, which minimized the impact that potential differences in difficulty and behavioral strategy could have had on our main conclusions.

Participants registered Intervening task responses with a keypress immediately after the last Intervening task stimulus. Thus, neural signatures of motor planning and execution may be

present in the EEG data near the end of the WM retention phase. However, we deemed this less detrimental than asking participants to hold their Intervening task responses until the end of the trial, as such a design would increase the WM load in trials with an Intervening task and complicate comparison with the no Intervening task conditions. Furthermore, any motor components in the EEG data should be present whenever there was an Intervening task, allowing for fair comparison between the different combinations of WM and Intervening tasks.

Trials were grouped into blocks of 20. Within each block, the WM and (when present) Intervening task conditions were held constant. At the start of each block, an instruction screen indicated the sensory modality and relevant domain (temporal or spatial) for both tasks in the upcoming block. Participants were allowed to take untimed breaks between blocks. Participants performed one block of each WM and Intervening task combination to be completed in that session before any conditions were repeated, and the same condition was not allowed to repeat in adjacent blocks. In total, participants performed 40 trials of each combination of WM task modality, WM task domain, and Intervening task condition.

Each complete dataset required three separate visits to the lab. The first session was reserved for consent, audiometric screening, and task practice. Participants practiced each variant of the WM and Intervening tasks in isolation until they understood the procedure, then performed three to five example trials of the full dual-task paradigm. Data collection for the actual experiment occurred in the two subsequent sessions, which were split based on the sensory modality of the WM tasks; auditory and visual WM tasks were performed on separate days. The order of these sessions was randomized and counterbalanced across participants.

Stimulus Details

For the visual WM tasks, 12 stimuli were arranged in a circle centered on the fixation cross and shown on a constant dark grey background (2.51 cd/m^2). Each stimulus was a square patch of visual noise, subtending 2.86° of visual angle and composed of a 30 x 30 grid of smaller squares. Each of these smaller squares was filled with a greyscale color between black and white, such that the average luminance across the patch was 5.85 cd/m^2. The angular spacing between each patch was 30°, and the entire stimulus circle subtended 21.59° of visual angle. To equate display luminance and structure across tasks, these visual stimuli remained present but static throughout the auditory WM and Intervening tasks. To employ these stimuli in the visual WM tasks, the luminance of each small square in a given patch could be resampled; this made the visual patch appear to jitter without changing the average luminance across the patch.

For the auditory WM tasks, each stimulus was a 50-ms tonal chord consisting of 3 harmonically unrelated complex tones (fundamental frequencies of 422, 563, and 670 Hz); each complex was made up of its first nine harmonics with equal amplitude. The same tone complex was used for all auditory stimuli in the WM tasks. For the Intervening tasks, which were always auditory, the stimuli were one of five pre-generated, 50-ms bursts of noise bandpass filtered between 100 and 10,000 Hz. Identical noise tokens were used for all three stimuli within each Intervening task sequence. Both types of stimuli were relatively broadband, thus ensuring they provided rich and robust spatial localization cues. All auditory stimuli were ramped on and off with a 5 ms cosine-squared ramp to avoid spectral splatter and transient onset and offset artifacts.

In the visual WM tasks, the first stimulus to change in the encoding and probe sequences was always the one at top-center (12 o'clock). Similarly, in all auditory WM sequences, the first sound was presented from the central loudspeaker. This was done to equate the number of items

stored in WM across the spatial and temporal WM tasks; with four stimuli in each sequence, there were three intervals to remember for the temporal tasks, and so the first stimulus location was held constant such that only three locations needed to be remembered for the spatial tasks.

Behavioral Data Analysis

The primary behavioral metrics in this study were error rates on the Intervening and WM tasks. Behavioral performance was statistically analyzed using logistic mixed effects regression models. For Intervening task performance, the model included fixed-effect terms for the WM task condition (AT, AS, VT, or VS) and the Intervening task type (AT or AS), as well as the interactions between these terms. Random-effects terms were included to capture participant-specific intercepts and slopes for both predictor variables. The model was structured as follows:

$$logit(Error\ Rate) \sim WM\ Condition * Int + (1 + WM\ Condition + Int \mid Participant)$$

A similar model was used to analyze WM recall errors. To facilitate separately examining effects of WM task modality (auditory or visual) and domain (temporal or spatial), the WM condition was expressed as two fixed-effect terms in this model:

$$logit(Error\ Rate) \sim Modality_{WM} * Domain_{WM} * Int$$
$$+ (1 + Modality_{WM} + Domain_{WM} + Int \mid Participant)$$

To investigate main effects and interactions at the group level, the coefficients from these models were fed into a two-way (Intervening task) or three-way (WM task) repeated measures ANOVA. P-values for model terms were estimated using the Satterthwaite approximation for degrees of freedom. Contrasts were treatment coded with baseline levels initially set as “auditory”

for WM modality, “temporal” for WM domain, and “none” for the Intervening task. These choices had no impact on the outcome of the group-level ANOVA. Pairwise post-hoc testing was conducted by cycling which level was considered “baseline” for each factor until a β-weight (and corresponding p-value) could be extracted for each necessary pair of levels.

We did not attempt to precisely equate task difficulty across conditions for individual participants, which makes it difficult to compare performance across Intervening task or WM task conditions. Instead, we assessed how performance on one task (keeping the condition of that task fixed) changed across conditions of the other task. For Intervening task performance, this meant examining performance on each separate Intervening task as a function of the modality and domain of the WM task; for WM task performance, we examined recall in each separate WM condition as a function of the Intervening task condition. All post-hoc tests were corrected for multiple comparisons using the Bonferroni-Holm method, using the total number of comparisons made in each analysis. For example, the four WM task conditions led to six comparisons within each Intervening task, but the number of comparisons for correction was set to the total across both Intervening tasks (12).

Pupillometry data collection and analysis

We measured task-evoked pupil dilations because they dynamically reflect the participant’s level of effort or arousal during task performance (Causse et al., 2016; Gilzenrat et al., 2012; Murphy et al., 2011; Winn et al., 2015). Prior to the start of each experimental session, eye position measurements were calibrated based on five fixation points (display center, ±20° azimuth on the horizontal plane, and ±10° elevation on the median plane). This calibration was validated at the center position prior to each trial, and at all five points at the start of each block.

Whenever eye position was offset from the calibration by an average of greater than 4°, the full five-point calibration was repeated. The experimenter monitored gaze position throughout the experiment; if at any point during the trial the participant's gaze deviated substantially from 0° in either azimuth or elevation, the experimenter immediately stepped in to ensure that the participant maintained center fixation moving forward.

A custom MATLAB analysis pipeline and wrapper GUI were used to prepare pupil data for statistical analysis. First, trials were split into baseline and trial windows, where the trial window spanned the start of the WM encoding sequence through the end of the WM retention phase. Next, blinks were automatically detected based on instantaneous position, velocity, and acceleration thresholds. An experimenter manually reviewed the data, and using the GUI, adjusted blink thresholds or manually marked additional blink segments as needed. Blinks and other marked segments of noisy data were replaced with a linear interpolation between the average of the three samples preceding and following the blink. When blinks occurred at the beginning of the trial window, a linear fit was made to the five samples following the blink, and this fit was back-projected through the blink segment. The opposite procedure was used for blinks falling at the end of the trial window. Trials in which more than 25% of the data was made up of rejected segments were automatically excluded from further analysis. When data from both eyes were available, the two traces were averaged. To produce the final output, the traces were concatenated, Z-scored, then split back into trials and trial windows. This procedure eschewed absolute pupil size measures in favor of values that were individually normalized for each participant, so only relative pupil diameter between conditions is interpreted.

Statistical testing for differences between conditions in the pupil time courses was carried out using non-parametric permutation tests. First, at each time point, the difference between two

conditions was assessed parametrically using a paired T-test across the individual participant average data. Next, we created a null distribution for each participant by randomly shuffling the condition labels and performing the same analysis for 2000 iterations. These randomly labeled samples were used to recompute paired T-tests at each time point, generating a null distribution of 2000 such values. Finally, at each time point, the significance level of the observed difference was determined by calculating the percentage of the null distribution with a T-value equal to or larger than the T-value from the actual data. Significant differences were only considered reliable if the p-value fell below 0.05 for a minimum of 15 consecutive samples.

EEG Data Analysis

EEG analyses were carried out using the FieldTrip package in MATLAB (Oostenveld et al., 2011). For event-related potential (ERP) analyses, EEG preprocessing comprised the following steps: read in the continuous data one channel at a time and immediately downsample to 256 Hz; re-reference the data to the average of two electrodes placed on the mastoids, bandpass filter between 0.5 and 20 Hz (zero-phase FIR filter, transition width of 0.2 Hz, order of 9274); manually identify and remove segments containing muscle artifacts; perform an independent components analysis (ICA) to project out blinks and saccadic eye movements; epoch the data from 100 ms before to 500 ms after each individual auditory or visual stimulus (timing differences between conditions precluded whole-trial averaging); reject any epochs in which the data exceeded a 100 μV peak-to-peak threshold; and baseline correct by subtracting off the mean of the first 100 ms of each epoch.

For time-frequency analyses, a similar preprocessing pipeline was used, but with a few key differences. First, the low-pass filter cutoff was raised to 80 Hz. Second, participant average ERPs

(recomputed with the new filter cutoffs) were subtracted from the data in the time domain immediately prior to epoching; this step served to minimize phase-locked evoked contributions to the time-frequency response. Subtracted ERPs were specific to each trial phase (i.e. encoding, retention, and probe), WM and Intervening task condition, and position in the stimulus sequence. Third, the data were split into whole-trial epochs, spanning the baseline period through the final stimulus in the probe sequence, instead of shorter individual stimulus epochs. The continuous Morlet wavelet transform (wavelet width of 5 cycles in 1 Hz steps) was used to obtain the power spectra of each trial. Prior to wavelet analysis, the signal was padded to avoid edge artifacts. This was done by copying the first and last 5 seconds of the epoch, reflecting each copy on the time axis, then appending them to the beginning and end of the signal, respectively. Finally, the data were split into the three key trial phases: WM encoding, WM retention/Intervening task, and WM probe.

We extracted time courses from the resulting power spectra at each channel in the theta (4-8 Hz) and alpha (8-13 Hz) frequency bands. To produce theta time courses, power estimates centered on 4, 5, 6, and 7 Hz (1 Hz bandwith) were averaged for each participant. Individual differences in the peak frequency of alpha oscillations are well-established (Klimesch et al., 1999). Thus, for alpha time courses, the individual alpha frequency for each participant was defined as the frequency at which the absolute value of the alpha change relative to baseline was maximal during WM retention (the direction of alpha change was found to flip based on WM modality in the present study). Alpha power time courses were reported as the power at this frequency averaged with the power 1 Hz above and below the individually defined alpha frequency.

Similar to the pupillometry data, EEG data were analyzed using non-parametric permutation testing, this time implemented using FieldTrip software. To control the multiple

comparisons problem with high-dimensional EEG data, we used cluster-based permutation testing to analyze ERPs and oscillatory power time courses (Maris & Oostenveld, 2007). Paired t-tests were first conducted at each time-channel data point. Those comparisons reaching a significance threshold (generally $p < 0.01$) were grouped into clusters of points contiguous in time and within 40 mm scalp distance of one another. The "mass" of this cluster was then computed as the sum of T-scores across all member points. Finally, this cluster mass was compared against a null distribution of clusters formed using the same analysis on 2000 random permutations of the condition labels.

Results

Intervening Task Performance

Performance accuracy on the Intervening tasks depended on the Intervening task condition (AT or AS) and its interaction with the type of information participants were holding in WM (AT, AS, VT, or VS; Fig. 2). These effects were analyzed using a logistic mixed-effects model, coefficients of which were then supplied to a Type-III ANOVA to determine the significance of effects at the group-level (see Materials and Methods). This ANOVA showed significant main effects of Intervening task ($\chi^2(1,20) = 32.7$, $p < 0.0001$) and WM task ($\chi^2(3,20) = 17.6$, $p = 5.29 \cdot 10^{-4}$), as well as a significant interaction between these factors ($\chi^2(3,20) = 24.1$, $p < 0.0001$). The main effect of Intervening task reflects overall lower error rates (better performance) on the AS compared to the AT Intervening task. The two loudspeakers used for the Intervening tasks were placed as close together as possible, but this spatial separation was nonetheless large enough for the AS Intervening task to be relatively easy for some participants. Given the overall performance difference between the two Intervening tasks, we examined each Intervening task

separately in subsequent pairwise post-hoc tests; differences in these tests could only be caused by interference based on the type of information participants were holding in WM.

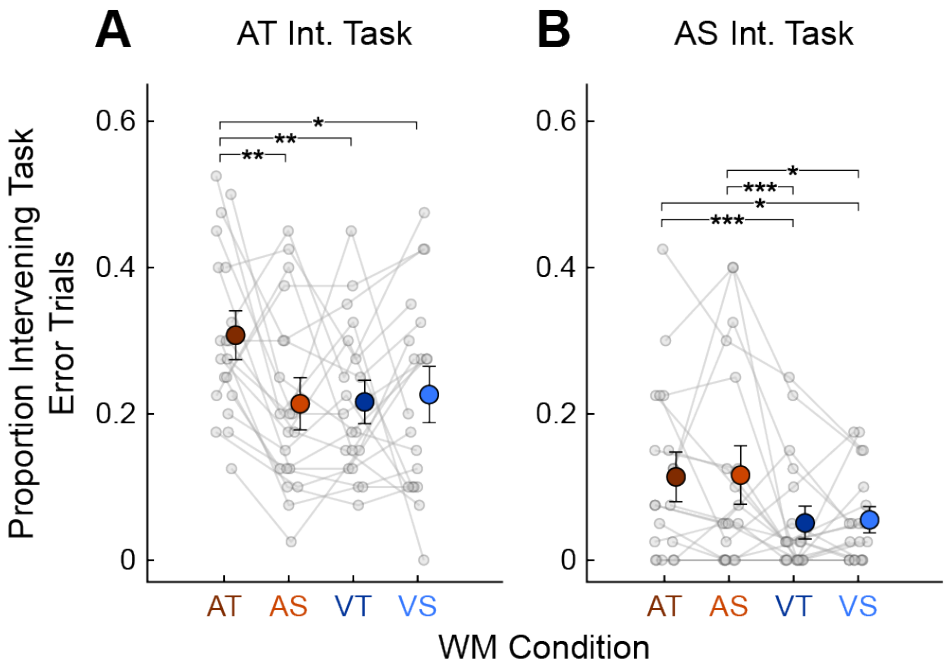

Figure 2: Intervening task error rates. Performance is shown for the AT (**A**) and AS (**B**) Intervening tasks. Grey points represent individual participants, large colored circles represent means, and error bars represent S.E.M. Chance performance is an error proportion of 0.5. Asterisks indicate significant differences between WM task levels in the mixed-effects model, examined separately within each Intervening task. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ after Bonferroni-Holm correction.

Participants made significantly more errors on the AT Intervening task (Fig. 2A) when the information being held in WM was also AT, as compared to all the other WM conditions (AT vs. AS, $p = 0.006$; AT vs. VT, $p = 0.003$; AT vs. VS, $p = 0.03$). Accuracy on the AS Intervening task (Fig. 2B) appeared to be modulated primarily by the sensory modality of the WM task. Participants made significantly more errors when both tasks were auditory than when the WM task was visual (AT vs. VT, $p = 1.76 \cdot 10^{-5}$; AT vs. VS, $p = 0.026$; AS vs. VT, $p = 5.50 \cdot 10^{-5}$; AS vs. VS, $p = 0.021$). We also hypothesized that errors on the AS Intervening task might increase when the WM task was VS, as both tasks shared spatial processing demands, but this did not appear to be the

462 case. However, the floor effect in AS Intervening task errors may have limited our ability to detect
463 additional interference effects based on WM domain.

464 Importantly, these patterns of Intervening task performance cannot be explained by
465 differences in task difficulty between the WM conditions. As will be shown below, participants
466 made relatively few recall errors in the AT WM condition in the absence of an Intervening task, at
467 least compared to the AS and VT WM conditions (see Fig. 3A below). If Intervening task
468 performance varied mainly with the difficulty of the WM task, one would therefore expect
469 relatively little interference in the AT WM condition, which in fact caused a relatively high degree
470 of interference in both Intervening tasks. Instead, the patterns here reflect relationships between
471 the types of processing demanded by the two tasks. Participants tended to make more errors on the
472 Intervening task when its sensory modality and (in the case of the AT Intervening task) its
473 information domain matched those of the information stored in WM.

474

475 *Working Memory Task Performance*

476 Participants were able to retrieve the information stored in WM at better-than-chance levels
477 across all combinations of WM modality, WM domain, and Intervening task. These factors
478 (including conditions with no Intervening task) served as fixed effects in a logistic mixed effects
479 model of WM task retrieval errors. A three-way ANOVA conducted on the resulting model
480 coefficients showed a significant interaction between WM task modality and domain ($\chi^2(1,20) =$
481 206.4, $p < 10^{-15}$), reflecting overall lower error rates on the AT and VS WM tasks (Fig. 3A). In
482 these conditions, the sensory modality was optimally suited for the information domain of the WM
483 task. The ANOVA also revealed a significant three-way interaction between WM modality, WM
484 domain, and Intervening task ($\chi^2(2,20) = 19.0$, $p = 7.66 \cdot 10^{-5}$), indicating different effects of the

Intervening tasks depending on the WM condition. These effects were explored via post-hoc testing conducted separately within each WM condition, so that differences in WM retrieval could be attributable solely to interference from the Intervening tasks.

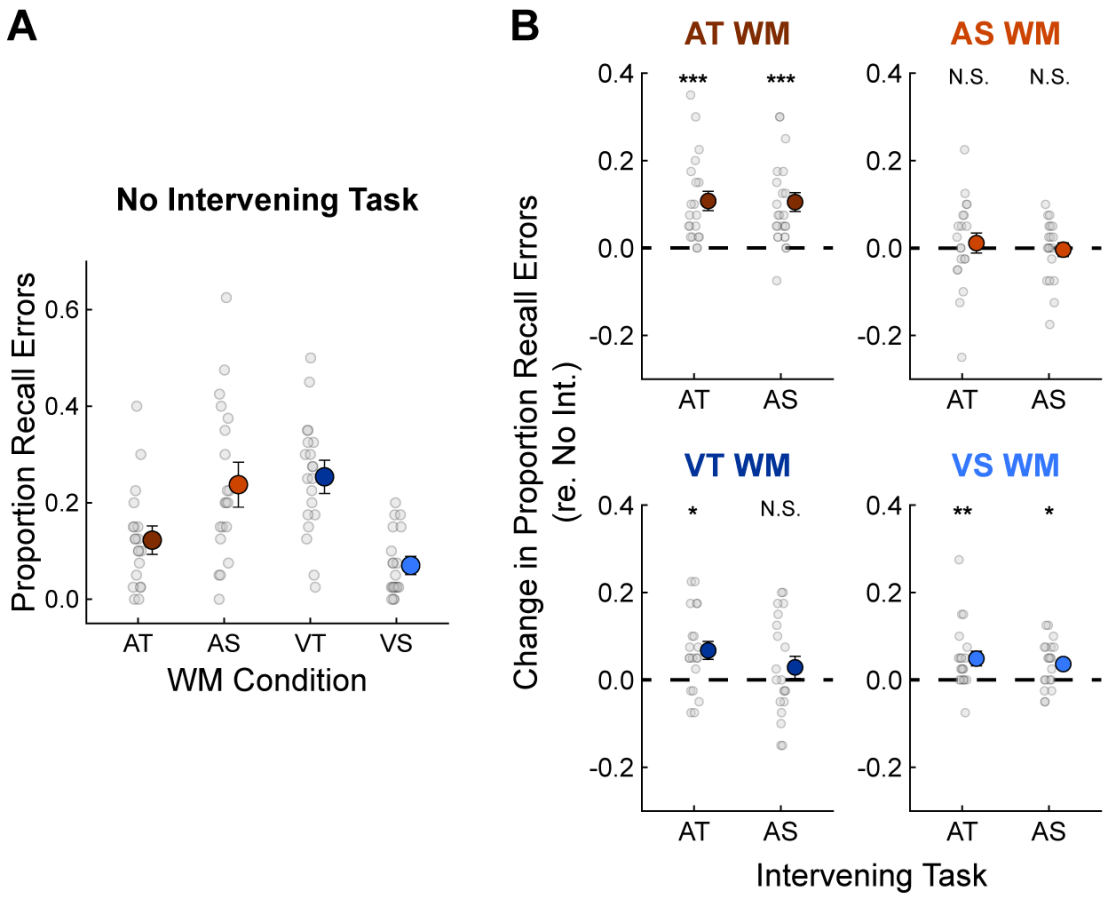

Figure 3: Working memory task performance. A, The proportion of trials on which the WM information was retrieved incorrectly is plotted for conditions without an Intervening task. Chance performance is at 0.5. **B**, The change in retrieval error rate is plotted for each WM and Intervening task combination relative to the no Intervening task conditions. The dashed line at 0 indicates no change in WM retrieval. Post-hoc comparisons were limited to differences between Intervening task conditions within each WM task. Asterisks indicate significant differences from the corresponding no Intervening task conditions (no significant differences were found between the two Intervening tasks). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ after Bonferroni-Holm correction.

To visualize effects of the Intervening tasks on recall accuracy, Fig. 3B shows the *difference* between WM error rates in each Intervening task condition and the corresponding condition with no Intervening task. When the WM task was AT (top-left panel), both auditory

Intervening tasks significantly impaired WM retrieval ($p < 10^{-6}$ for both). When the WM task was AS, on the other hand, the Intervening tasks had no detectable impact on WM retrieval (top-right panel). We suspect that holding AT information in WM relied heavily on the auditory-biased WM network, leading to memory interference from the auditory Intervening tasks. In the AS WM task, on the other hand, sound locations to be remembered may have been, at least in part, mapped into a representation in the visual-spatial WM network, helping to protect this information from auditory interference (see Discussion).

Recall accuracy was relatively poor in the AS WM condition, which may have impaired our ability to detect effects of the Intervening tasks in this condition, as closer-to-chance performance left less room for performance to be impaired by the Intervening tasks (Fig. 3A). To explore this possibility, we built a linear model of the effects of the Intervening tasks (the difference in recall performance between conditions with and without an Intervening task) as a function of performance in the no Intervening task condition. There was no statistically significant relationship when the Intervening task was AS ($p = 0.31$); thus, the degree of interference from the Intervening task did not depend on baseline AS WM performance. For the AT Intervening task, the effect was marginally significant ($F(1,18) = 4.71$, $p = 0.044$), but this effect hinged on the one participant who performed at worse-than-chance levels in the No Intervening task condition; without this participant, the effect disappeared ($F(1,17) = 0.54$, $p = 0.47$). Thus, we conclude that the auditory Intervening tasks had minimal effect on participants' ability to recall AS information from WM.

In some cases, visual WM retrieval was also modestly impaired by the auditory Intervening tasks. A significant increase in retrieval errors on the VT WM task was detected when the Intervening task was also temporal ($p = 0.042$; bottom-left panel of Fig. 3B). This might reflect

domain-based WM interference with both tasks drawing on temporal processing resources, although no significant difference was found between the AT and AS Intervening task conditions. In the VS WM condition, both Intervening tasks increased retrieval errors (p = 0.003 for AT Intervening task, p = 0.044 for AS; bottom-right panel).

Pupil Dilations

Both encoding information in WM and performing the auditory Intervening tasks led to reliable pupil dilations, shown in Fig. 4. In the absence of an Intervening task, pupil diameter increased throughout presentation of the stimuli to be encoded in WM (from −2 to 0 sec on the time axis), peaked about 700 ms after the offset of the final stimulus in the WM encoding phase, and then gradually declined throughout WM retention (Fig. 4A, left panel). Pupil size was significantly larger during the encoding phase of the AS WM condition compared to any of the other WM conditions.

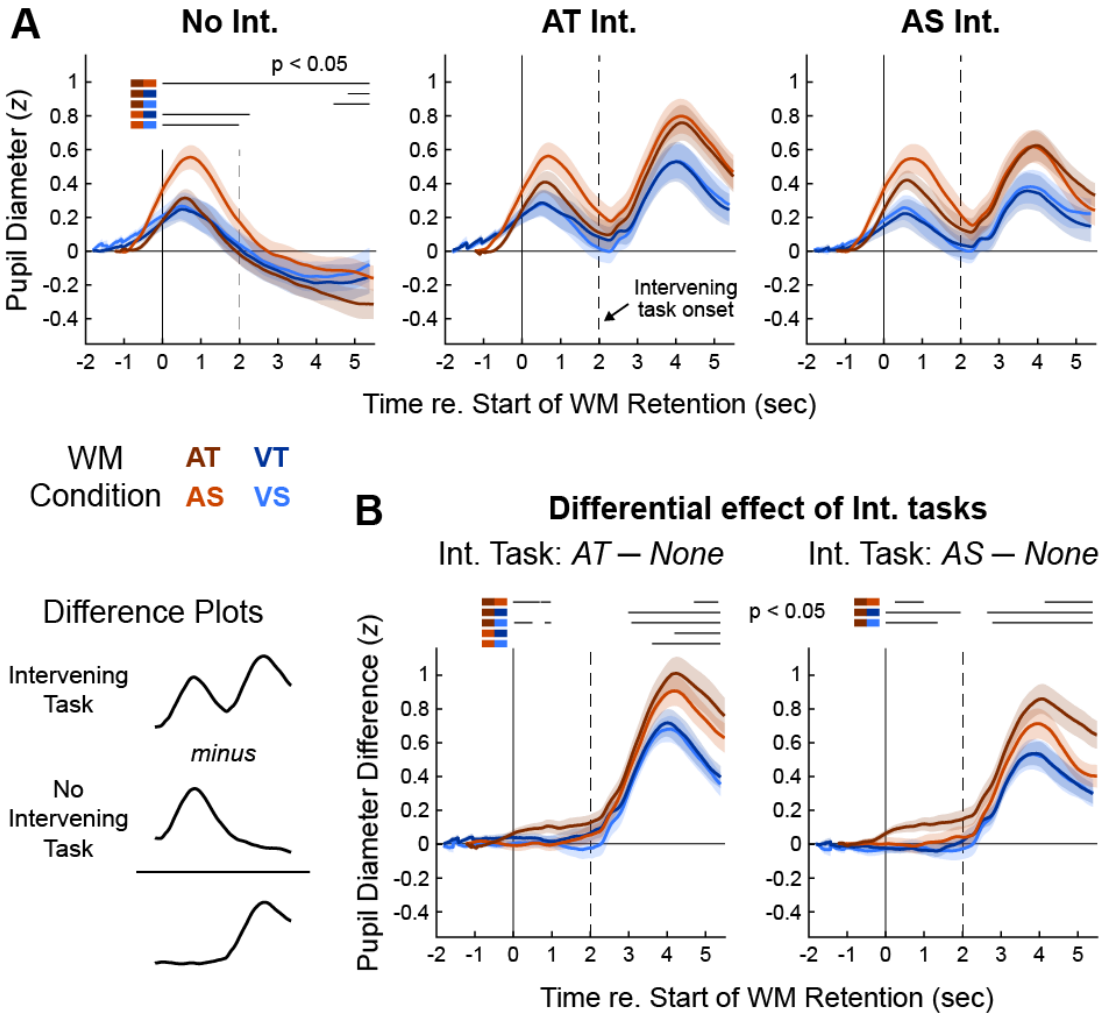

Figure 4: Pupillometry time courses. A, Grand average Z-scored pupil responses are shown in the No Intervening task condition (left) and elicited by physically identical stimuli in the AT (middle) and AS (right) Intervening tasks. Solid vertical lines represent the end of WM encoding and the start of WM retention, while dashed vertical lines indicate the onset of the Intervening task when present. Error clouds represent S.E.M. Horizontal black lines above the traces represent time regions of significant difference in permutation testing, with the two WM conditions being compared indicated by colors next to each significance line. **B**, The difference between pupil responses elicited in conditions with an Intervening task and responses in the corresponding WM conditions with no Intervening task.

There were also large secondary pupil dilations in conditions with an Intervening task, which started shortly after Intervening task onset but peaked around 4s into WM retention, approximately 1s after the Intervening task behavioral response. The amplitudes of these pupil responses were modulated by the information participants were holding in WM (Fig. 4A, middle

551 and right panels). To better isolate the impact of the Intervening tasks on pupil diameter, we
552 subtracted corresponding pupil traces with no Intervening task from each pupil trace with an
553 Intervening task (Fig. 4B). Statistical differences between the pupil responses were assessed using
554 non-parametric permutation testing, restricted to a time window spanning the WM retention phase
555 (time zero and later in Fig. 4). We first averaged the differential pupil traces across the four WM
556 conditions and tested for a difference between the two Intervening tasks (not shown). This
557 confirmed that traces elicited by the AT Intervening task were larger overall than those elicited by
558 the AS Intervening task ($p < 0.05$ for all time points between 3.7 sec after WM retention onset and
559 the end of the WM retention window), matching the higher behavioral error rates on the AT
560 Intervening task.

561 Next, we compared the differential pupil traces between the four WM conditions within
562 each Intervening task (Fig. 4B), similar to our treatment of the Intervening task behavioral data.
563 Differential pupil dilations elicited by both Intervening tasks were larger when the information
564 held in WM was auditory than when it was visual, mirroring the modality-driven interference
565 effects observed in the behavioral data. The largest differential pupil dilation occurred when the
566 WM task was AT—significantly larger than when the WM task was AS—for both Intervening
567 tasks. In the AT WM condition, the residual pupil traces were also elevated between 0 and 2
568 seconds from the start of WM retention (prior to the onset of the Intervening task), indicating a
569 preparatory increase in effort in this WM condition. Combined, these results suggest that
570 participants increased task effort both before and during the Intervening task when the information
571 held in WM was AT, as both auditory Intervening tasks strongly interfered with recall in this WM
572 condition. In the visual WM conditions, pupil dilations were insensitive to whether the WM task
573 was temporal or spatial. This result is similar to the Intervening task error rates and indicates that

interference was generally low when the two tasks were presented in different sensory modalities, even if they were matched in information domain.

We also analyzed average pupil size during the pre-trial baseline period, primarily to assess whether our baseline correction procedure may have biased the subsequent pupil dilation patterns. We conducted a separate two-way ANOVA for each Intervening task condition, with factors of WM modality and WM domain. In the AT Intervening task ANOVA, there was a significant main effect of WM modality ($F(1,19) = 9.89$, $p = 0.005$, partial-$\eta^2 = 0.045$), with baseline pupil size tending to be smaller in the auditory than the visual WM conditions (Fig. 5). This is an intriguing result, as the AT Intervening task was relatively difficult and tended to interfere with auditory (particularly AT) WM. Smaller pupil size in the auditory WM conditions might indicate fatigue during these challenging blocks (Morad et al., 2000). However, there were no significant baseline pupil size effects in the AS Intervening task ANOVA (nor the no Intervening task ANOVA). Since the pattern of task-evoked pupil dilations was not qualitatively different between the AT and AS Intervening task conditions, it is unlikely that baseline correction biased the main results in any meaningful way.

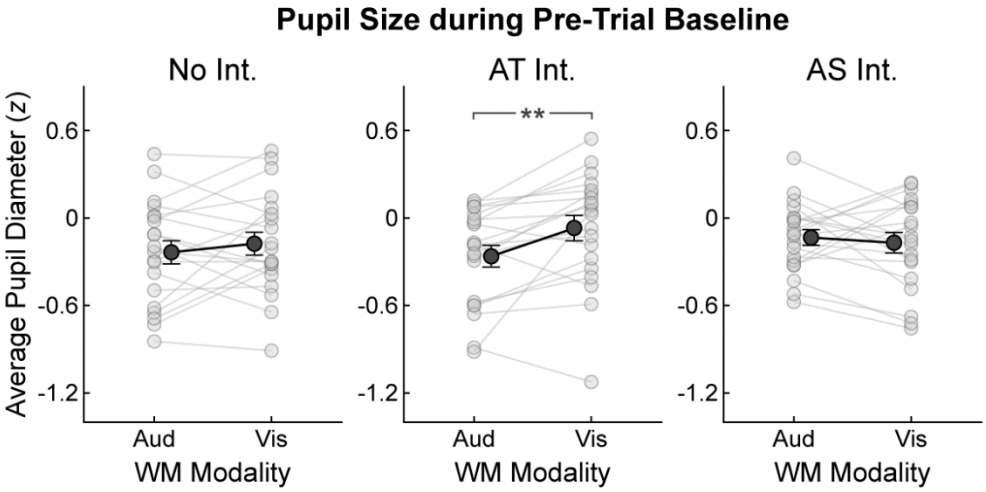

Figure 5: Baseline pupil size. Z-scored pupil size, averaged across the pre-trial baseline window, is shown for each Intervening task condition and WM task modality. Data are collapsed across WM domain, as this factor had no statistical influence on baseline pupil size. Light grey points and lines represent individual participants, dark grey points represent grand averages, and error bars represent S.E.M. ** = $p < 0.01$.

Event-related potential (ERP) amplitudes

While the behavioral and pupillometry data revealed consistent effects of the modality and domain of the WM task, the domain of the Intervening task had relatively little impact on signatures of dual-task interference. The AT and AS Intervening tasks did require different types of auditory processing however, which could affect patterns of dual-task interference at the neural level. We next investigated this possibility in the EEG data.

We first examined event-related potentials (ERPs) elicited by each stimulus event in the WM and Intervening tasks (Fig. 6). ERPs elicited by stimuli in the encoding and probe sequences of the WM task were not systematically modulated by the Intervening tasks. However, the magnitude of the ERPs elicited by the Intervening task stimuli—particularly the first sound in the Intervening task sequence (highlighted in Fig. 6)—varied depending on the information participants were holding in WM. The Intervening task auditory ERPs had a somewhat atypical

morphology, including a weak N1 and a strong component in the P2 time range (compare the Intervening task ERPs to auditory ERPs elicited by the WM task stimuli). For convenience, we will refer to this positive component as the "P2" and return to its interpretation in the Discussion.

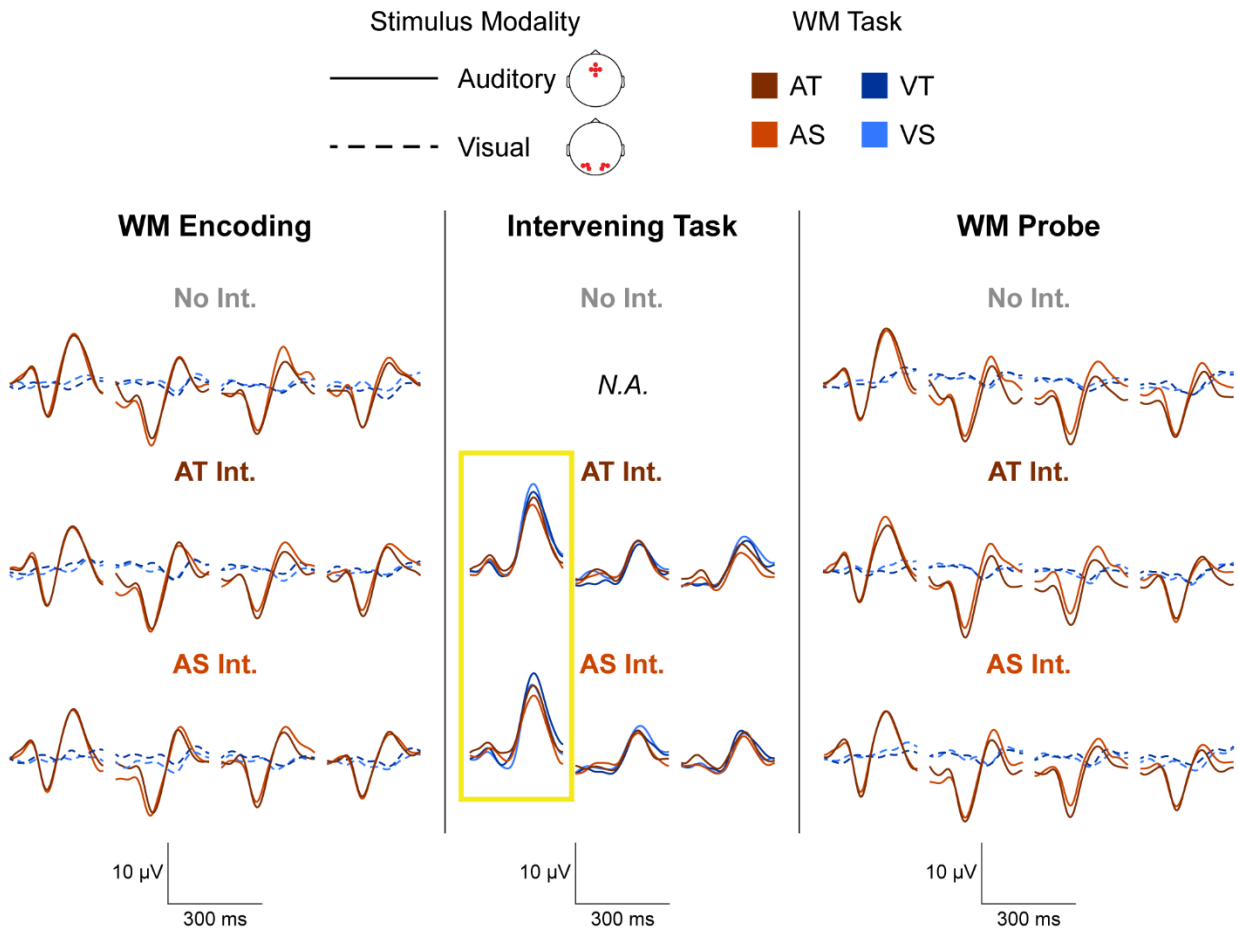

Figure 6: Event-related potentials. Grand-average ERPs are shown for each combination of WM task (colors), Intervening task (rows), and position in the stimulus sequence. The modality of stimulus presentation is shown by solid versus dashed lines. Intervening task onset ERPs, which were subsequently analyzed in greater detail, are highlighted by the yellow box. The top-left legend shows the channels averaged to produce ERPs for each stimulus modality: on the standard 10-20 layout, these were channels Fz, AFz, Cz, F1, and F2 for auditory stimuli and O1, O2, PO3, PO4, PO7, PO8 for visual stimuli. Axes below the data in each trial phase are shown for scale.

Both the modality and domain of the information stored in WM affected the amplitudes of the P2 component (Fig. 7, panels A and B). To quantify these effects, P2 amplitudes were calculated for each participant as the average of the ERP waveform between 170 and 230 ms post-

stimulus across a cluster of fronto-central electrode sites (Fz, FCz, Cz, FC1, and FC2 on the standard 10-20 layout). These amplitudes were then submitted to a three-way ANOVA with explanatory factors of WM modality, WM domain, and Intervening task. This analysis revealed a significant main effect of WM modality ($F(1,19) = 7.27$, $p = 0.014$, $\eta^2 = 0.038$) and a significant interaction between WM domain and Intervening task ($F(1,19) = 10.83$, $p = 0.004$, $\eta^2 = 0.005$).

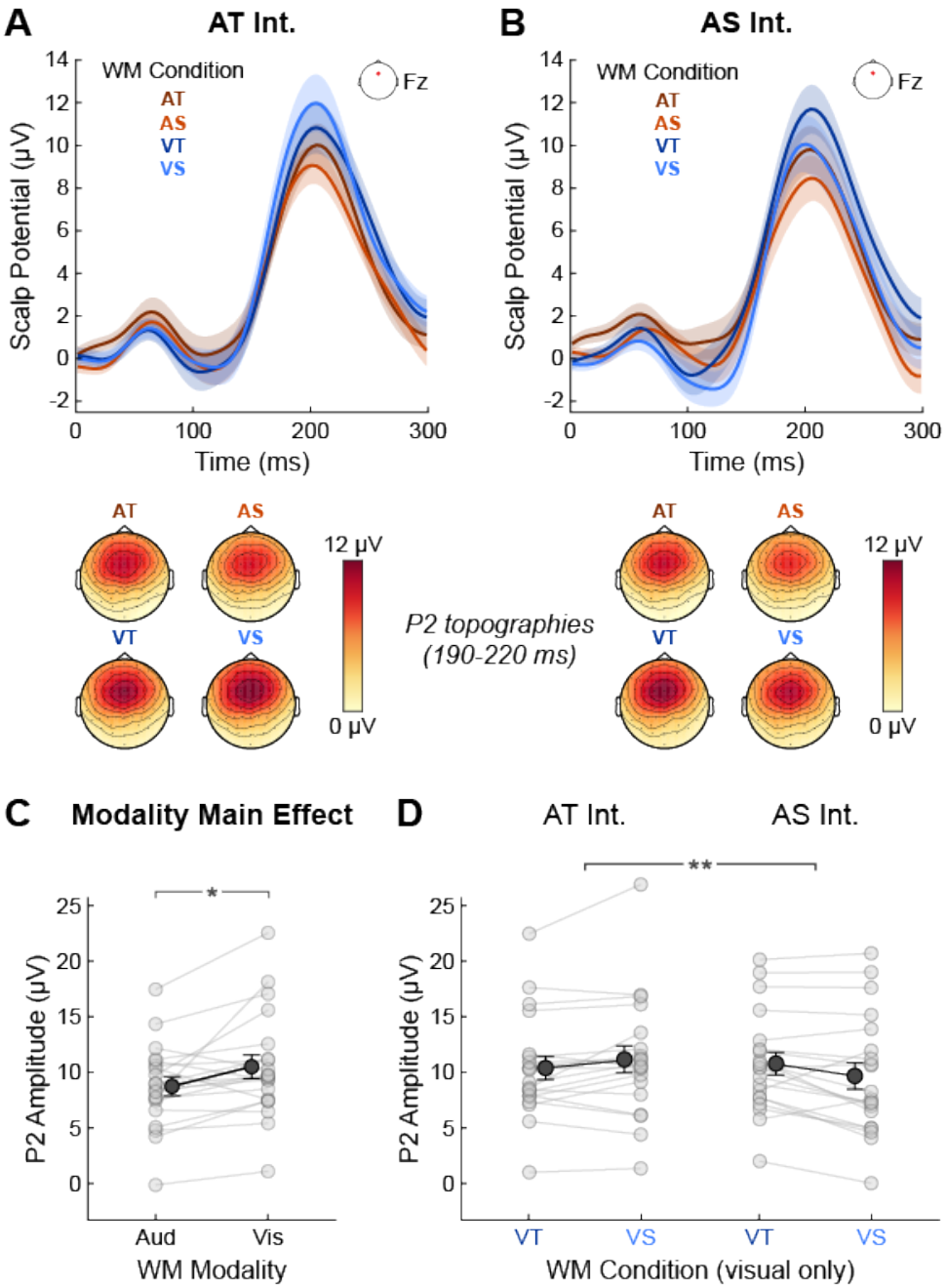

Figure 7: Intervening task onset ERPs. A, Grand average ERPs elicited by the first sound in the AT Intervening task at a fronto-central electrode site (Fz). Color represents the WM task, and error clouds represent S.E.M. Scalp topographies of the P2 responses are shown below the traces. **B**, Same as A, but for the AS Intervening task. The ERPs shown in A and B are those highlighted in Fig. 6. **C**, P2 peaks in the auditory and visual WM conditions, averaged across WM task domain and Intervening task. **D**, P2 peaks in the visual WM conditions to illustrate the interaction between WM domain and Intervening task.

630

The main effect of WM modality reflects the fact that P2 amplitudes elicited by both auditory Intervening tasks were larger when visual rather than auditory information was stored in WM (Fig. 7C). Although the three-way interaction involving WM task modality did not reach statistical significance, the significant interaction appears to have been driven predominantly by the visual WM conditions. We therefore split the P2 peak data into auditory and visual WM conditions, then conducted separate 2-way ANOVAs with factors of WM domain and Intervening task.

In the visual WM conditions, the interaction between WM domain and Intervening task was significant ($F(1,19) = 10.77$, $\eta^2 = 0.009$, $p = 0.004$; beneath a Bonferroni-Holm adjusted alpha criterion to account for the two additional ANOVAs). Specifically, ERPs elicited by the AT Intervening task were slightly larger when visual-*spatial* information was stored in WM, whereas ERPs elicited by the AS Intervening task were larger when visual-*temporal* information was stored in WM (Fig. 7D). This provides evidence that P2 components of the ERPs were reduced when the two tasks required processing in the same information domain. However, in the auditory WM ANOVA, there was a significant effect of WM domain ($F(1,19) = 5.21$, $p = 0.034$, $\eta^2 = 0.015$) but no interaction with Intervening task; although the significance of the WM effect did not survive Bonferroni-Holm adjustment of the alpha criterion, the trend was for ERPs elicited by either Intervening task to be larger when AT information was stored in WM than when the WM task was AS.

Alpha-band oscillatory activity

Changes in oscillatory activity in the alpha band (8-13 Hz) have been linked to maintaining information in WM (Bastiaansen et al., 2002; Obleser et al., 2012). We investigated whether this

neural signature of WM maintenance was affected by the presence of an Intervening task, as well as the degree of similarity in sensory modality and information domain between the Intervening and WM tasks. Figure 8A shows the power spectrum during the pre-trial baseline period at a posterior electrode site (Pz), with a strong peak in the alpha range, validating that this oscillatory signature occurred in the present data. An anterior electrode site is also shown, with expected oscillatory power in the 4-7 Hz range (frontal theta oscillations).

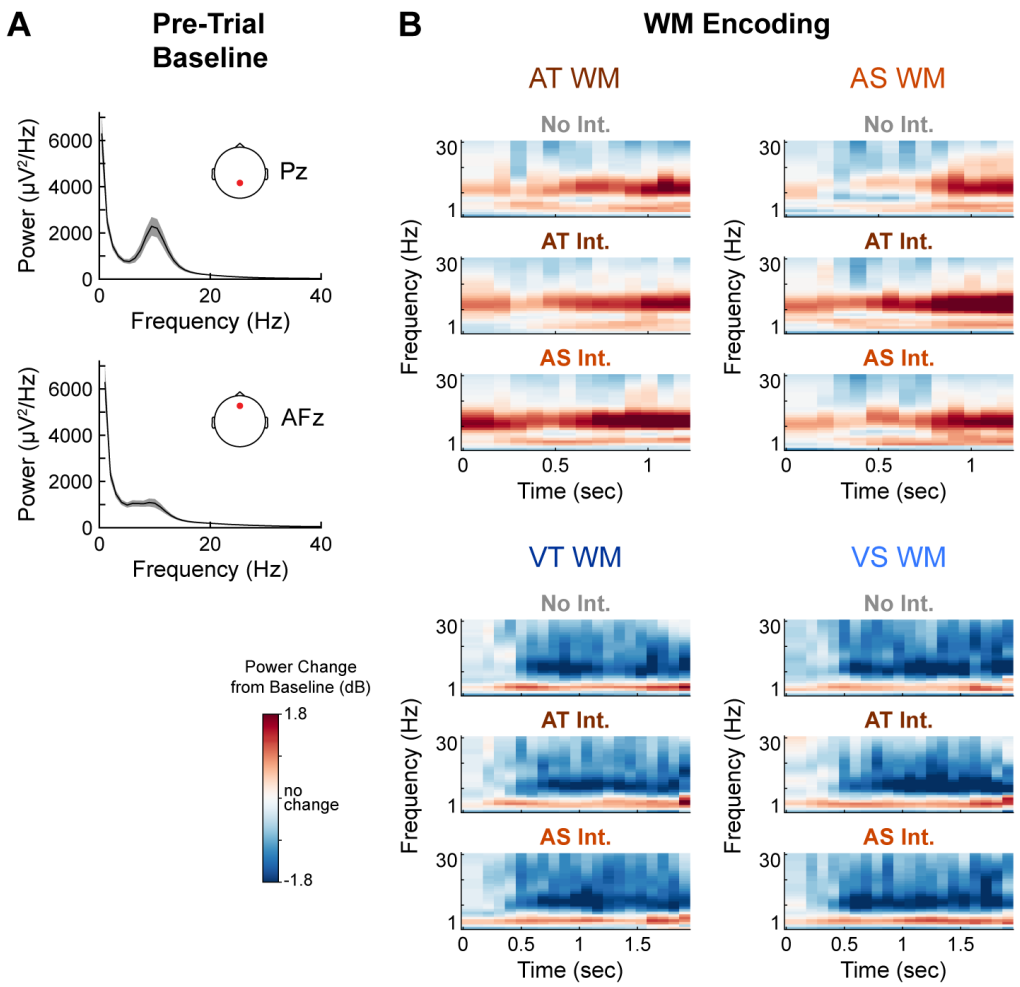

Figure 8: Oscillatory power in the baseline and during WM encoding. A, Grand average power spectra during the pre-trial baseline, plotted at a posterior and an anterior electrode site. Error clouds represent S.E.M. **B**, Grand average oscillatory power during WM encoding in the AT, AS, VT, and VS WM conditions. Data are shown as change relative to the pre-trial baseline, in dB. Because the length of the encoding window differed based on inter-stimulus intervals, responses of different lengths were aligned to the end of the encoding phase prior to averaging. Similarly, because the long interval was longer for visual WM than auditory WM, the time bases differ between the top and bottom panels.

669

670 During WM encoding, alpha power steadily increased relative to baseline in the auditory
671 WM conditions and decreased relative to baseline in the visual WM conditions (Fig. 8B). These
672 gross patterns of alpha activity persisted into the WM retention window. Preliminary analysis of
673 alpha power across the retention window using cluster-based permutation testing revealed no
674 significant effects of WM task domain, so data were collapsed across this factor and subsequently
675 analyzed based on WM task modality and the Intervening task.

676 In the auditory WM tasks, alpha power remained elevated relative to baseline throughout
677 the memory retention window when no Intervening task was present (Fig. 9A, top panel). The
678 onset of both the AT and AS Intervening tasks suppressed these ongoing alpha oscillations (middle
679 and bottom panels). We examined this effect by conducting cluster-based permutation tests on the
680 alpha power time courses in two time regions of interest: one spanning the Intervening task itself
681 (2 to 3.5 seconds into WM retention, including the Intervening task behavioral response), and
682 another from 3.5 to 5 sec, during which participants needed to start retrieving the information
683 stored in WM in anticipation of the WM task probe sequence. During both of the Intervening tasks
684 (the earlier window), alpha power was significantly reduced relative to the no Intervening task
685 condition (AT Intervening task, $p = 0.005$; AS Intervening task, $p = 0.003$; Fig 9B). With the
686 statistical threshold for generating the initial clusters set at $p = 0.01$, no differences were found
687 between alpha power in the two Intervening tasks. However, relaxing this threshold to $p = 0.05$
688 (not shown) revealed an additional late cluster showing an alpha power difference between the AT
689 and AS Intervening tasks (significant between 3.14 and 3.45 sec; $p = 0.026$). This suggests that
690 the alpha power reduction lasted longer when the Intervening task was AT than when it was AS,
691 probably because participants could sometimes make their perceptual judgment earlier in the AS

Intervening task by comparing the first sound location to the second (without using the third). No significant clusters were found in the later (3.5-5 sec) time window in the auditory WM conditions.

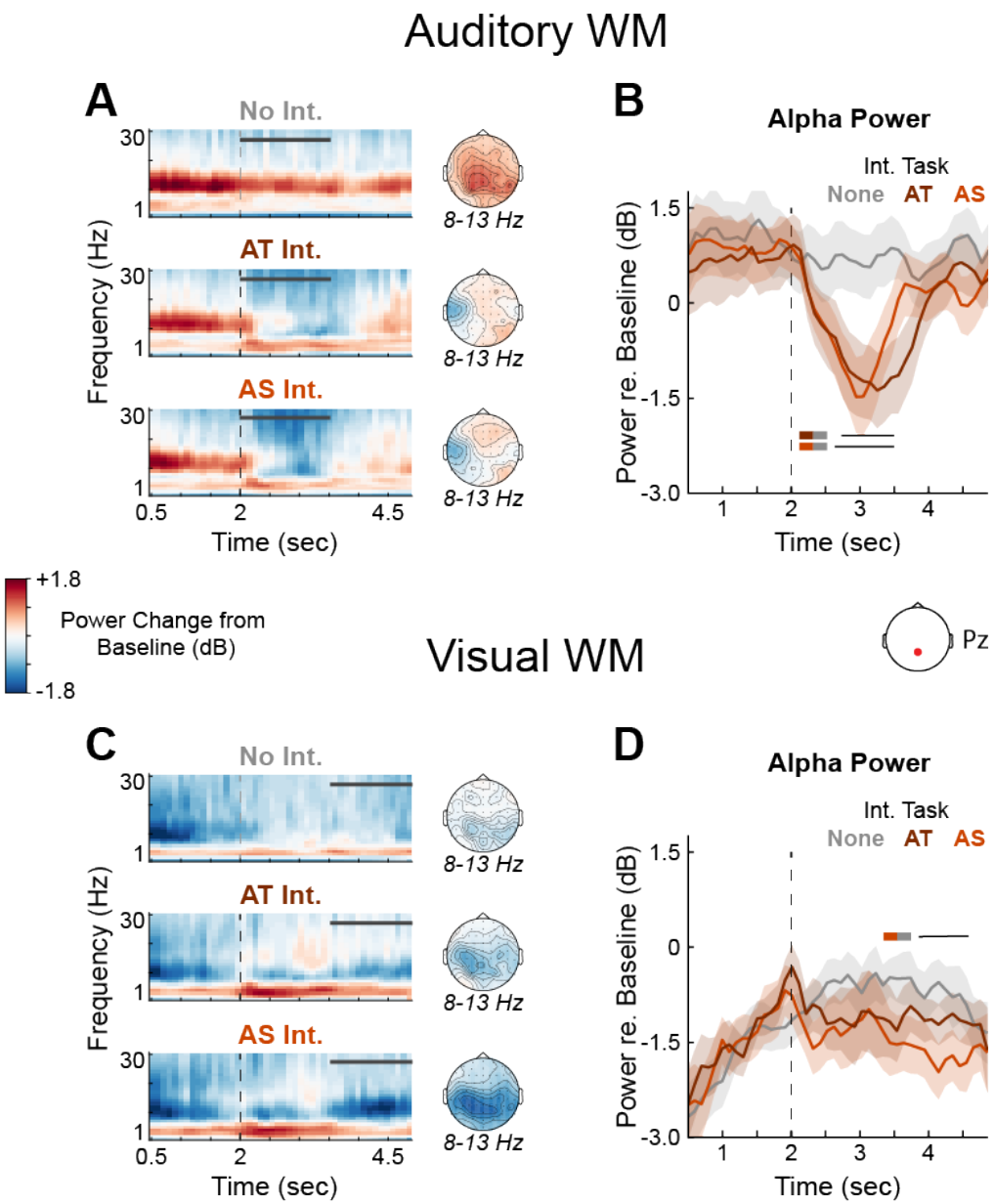

Figure 9: Alpha power during WM retention. Grand-average time-frequency responses are shown at a parietal electrode site (Pz) during the memory retention window for each of the three Intervening tasks in auditory (**A**) and visual (**C**) WM conditions. Responses are averaged across WM domain and shown as dB change relative to the average pre-trial baseline period across conditions. Vertical dashed lines represent Intervening task onset when present, and horizontal grey bars represent the time regions in which significant effects were found in cluster-based permutation testing. The first and last 500 ms of the retention window were excluded to limit power contributions from responses evoked by the WM task stimuli. B, Grand average alpha power time courses are also shown for the auditory (**B**) and visual **(D)** WM conditions. Power time courses for each participant were calculated at their peak alpha frequency ±1 Hz. Error clouds represent S.E.M**.** Black horizontal bars indicate the temporal extent of significant clusters in cluster-based permutation testing for the comparisons indicated by colored boxes to the left.

In the visual WM conditions without an Intervening task, alpha power gradually increased from the suppressed levels during WM encoding to near-baseline levels during WM retention (Fig. 9C, top panel). On trials with an Intervening task, alpha power become suppressed again after the conclusion of the Intervening task (bottom panels). We examined the alpha power time courses using cluster-based permutation testing restricted to the same two time windows as in the auditory WM conditions. Unlike the auditory conditions, no significant differences were observed during the Intervening tasks. However, in the later window, when participants needed to transition from the Intervening task to WM recall, alpha power was significantly reduced when the Intervening task was AS relative to when there was no Intervening task ($p = 0.011$; Fig. 9D). Alpha power did not differ either between the AT and no Intervening task conditions or between the AT and AS Intervening tasks.

Finally, we examined frontal theta band (4-7 Hz) oscillatory activity, as increased theta power has been implicated in paradigms that require task-switching (Cunillera et al., 2012; López et al., 2019; Sauseng et al., 2006). The presence of either Intervening task caused an increase in frontal theta power throughout the memory retention window (not shown). In all but one of the conditions with an Intervening task, cluster-based permutation tests revealed at least one time-channel cluster in which theta power was significantly elevated relative to the corresponding no Intervening task conditions ($p < 0.05$ for all clusters). In the remaining condition (AT WM, AS Intervening task), a similar increase in theta power trended toward significance ($p = 0.053$). Although ERPs can manifest as activity in this frequency range as well, the increase in theta power was observed both preceding and well after the Intervening task, meaning it cannot easily be explained by evoked responses (grand average, condition-specific ERPs were also subtracted from the time series prior to time-frequency analyses). The broad presence of increased theta power

across conditions with an Intervening task suggests that switching from WM encoding to perceptual processing drove this neural signature, but it generally did not differ based on the condition of either task.

Discussion

In this study, we investigated interference between perceptual processing and WM maintenance while varying similarity in sensory modality and information domain. Convergent evidence from behavior and pupillometry (indexing cognitive effort) showed that the highest degree of interference occurred when the two tasks drew upon shared cognitive resources. Specifically, when both tasks were auditory-temporal (AT), error rates on the Intervening task were highest, WM retrieval was poorest, and pupil dilations were largest. In contrast, when the auditory Intervening tasks were paired with a visual (visual-temporal, VT or visual-spatial, VS) WM condition, behavioral interference was relatively low and pupil size was smallest. In these conditions with reduced competition for shared neural resources, subtler patterns of interference related to the task domain were nonetheless present in the EEG data. Results from behavior and pupillometry indicated an intermediate level of interference from the auditory Intervening tasks when the WM task was auditory-spatial (AS); this condition will receive special attention below, as we expected maintaining AS information in WM to draw on both auditory- and visual-biased control networks, leading to more nuanced patterns of interference.

Interference patterns support modality and information domain specializations of lateral frontal control networks

The design of this experiment was inspired by recently characterized attention and WM networks involving human lateral frontal cortex (LFC). Subregions of the LFC preferentially contribute to auditory or visual processing, but are also recruited by stimuli in the non-preferred modality depending on whether the task is temporal or spatial. In this study, our goal was to examine the influence of sensory modality and information domain on dual-task interference, informed by the response profiles of these networks. We did not use neuroimaging methods (such as fMRI) that would allow us to directly relate dual-task interference to activity in subregions of the LFC, but our WM task conditions were derived from tasks shown to recruit these networks in fMRI experiments (Michalka et al., 2015). Thus, we are confident that the AT WM condition strongly engaged the auditory-biased network, the VS WM condition strongly engaged the visual-biased network, and the AS and VT WM conditions recruited both networks. However, recruitment of these LFC networks is also asymmetric; auditory processing activates the visual-biased network to a greater extent than visual processing activates the auditory-biased network (Noyce et al., 2017, 2022). The AS WM condition and AS Intervening task are therefore of particular interest in the present study, as we expected that auditory information in these tasks might ultimately be mapped to, and maintained in, the visual-biased network. One prediction from this is that perceptual processing that taxes the auditory-biased network should be less affected by holding AS information than holding AT information in WM, as the AS information can be represented (at least partially) in the complementary visual-spatial network. In line with this prediction, participants made fewer errors on the AT Intervening task (which we expected to load exclusively onto the auditory-biased network) when the WM task was AS than when it was AT.

763 The AS Intervening task, on the other hand, may have recruited both networks to some extent,
764 leading to similar levels of behavioral interference from a WM condition that overloaded the
765 auditory-biased network (AT WM) and one that increased the overall distributed load across the
766 two networks (AS WM).

767 We initially expected that interference effects between the two tasks would be
768 symmetrical—that is, that effects of the Intervening task on WM recall would be similar to effects
769 of the WM task on Intervening task performance. However, whereas the WM conditions
770 differentially affected Intervening task performance, the two Intervening tasks interfered with WM
771 recall roughly equally. This likely relates to important differences in processing demands between
772 the AS WM task and the AS Intervening task. To perform the AS WM task, participants needed
773 to remember absolute auditory stimulus locations mapped onto physical space. This may have
774 placed more demand on the visual-biased LFC network than the AS Intervening task, which
775 required only an immediate, relative judgment about auditory spatial positions within each trial,
776 and thus did not need to be stored in memory. While the AS Intervening task certainly involved a
777 form of spatial processing that could interfere with other spatial tasks, it seems likely that it
778 primarily taxed the auditory-biased LFC network, similar to the AT Intervening task. This would
779 explain why both Intervening tasks strongly interfered with AT, but not AS, WM recall, and why
780 pupil dilations elicited by both Intervening tasks showed the highest effort with AT information in
781 WM. An important (if unintended) point to draw from this is that two tasks with ostensibly similar
782 processing demands could in actuality draw upon different neural resources; in this case, not all
783 spatial tasks are created equal.

784 Another noteworthy aspect of the results presented here is that the processing costs of
785 loading onto a single network were greater than the costs of switching between networks. This

786 outcome is not entirely obvious, as task-switching also incurs a processing cost (Arnell & Jolicœur,
787 1999; Chun & Potter, 2001; Hsieh & Allport, 1994; Meiran et al., 2000). Our auditory Intervening
788 tasks did modestly impair visual WM retrieval, consistent with a behavioral cost of switching
789 between the complementary networks. However, this effect was relatively weak compared to the
790 behavioral (and physiological) costs when the tasks loaded onto shared network resources. Thus,
791 the ability to simultaneously leverage complementary attention and WM networks appears to
792 confer a processing benefit, which may be especially relevant in settings with simultaneous inputs
793 from multiple sensory modalities.

794 In this study, we only used auditory Intervening tasks to limit the influence of visual stimuli
795 on the pupil response. The lack of visual Intervening tasks resulted in an upper limit on how much
796 load could be placed on the visual-biased control network, as there was no condition involving
797 competing visual tasks. Future studies (even if they are only behavioral) should implement visual
798 Intervening tasks in a paradigm like this one to examine whether dual-task conditions that load
799 exclusively on the visual-biased network (i.e., two visual-spatial tasks) also result in elevated
800 interference. Domain-based patterns of interference might be more muted in such an experiment,
801 however, as recruitment of the auditory-biased LFC network by visual-temporal tasks is relatively
802 modest (Noyce et al., 2017).

803

804 *Unique contributions from pupillometry*

805 The pupillometry data in this study revealed details about the magnitude and time course
806 of interference that could not be discerned from behavioral responses alone. For instance, the AS
807 Intervening task was easy enough that a majority of participants did not make errors on it in any
808 WM condition. However, substantial differences in the amplitude of pupil dilations elicited by this

809 task suggest that AS perceptual processing and WM maintenance did cause mutual interference,
810 which participants overcame to achieve near-ceiling performance. Indeed, a major benefit of
811 pupillometry is that it can reflect differences in the effort required to reach a certain performance
812 level in the absence of behavioral differences (e.g., McGarrigle et al., 2017; Ohlenforst et al., 2017;
813 Winn & Teece, 2021). Pupil responses have also provided useful insights into the temporal
814 dynamics of effort allocation in cognitive processes that unfold through time, such as decision
815 making (Satterthwaite et al., 2007) and handling talker variability (Lim et al., 2021). In this study,
816 when the WM task was AT, pupil size increased prior to the onset of the Intervening task stimuli.
817 This suggests a preemptive modulation of effort, perhaps reflecting the need to preserve the WM
818 trace in anticipation of auditory interference.

819 The main pupil responses of interest in this study were those elicited by Intervening tasks
820 under different types of WM load. These pupil dilations started shortly after Intervening task onset,
821 but pupil size peaked after the Intervening task behavioral response and remained elevated
822 throughout WM retention. Thus, these responses likely represent the combined effects of
823 performing the Intervening tasks under WM load and recalling information from WM after
824 auditory interference. This could explain why they are in accord with some aspects of both the
825 WM task and Intervening task behavioral data. Consistent with WM recall, pupil responses were
826 largest in the AT WM condition, and the pattern across WM conditions was similar between the
827 two Intervening tasks. Consistent with Intervening task performance, pupil dilations were
828 intermediate in the AS WM condition (when analyzed as the difference from the no Intervening
829 task conditions), hinting that recruitment of the visual-biased network to store the AS WM
830 information reduced the effort required to overcome auditory interference.

Many studies have shown that pupil size scales with task difficulty (Bijleveld et al., 2009; Kahneman & Beatty, 1966; Porter et al., 2007; van der Meer et al., 2010; Zhao et al., 2019), at least up to the point that the task becomes so difficult that participants disengage from it (Granholm et al., 1996). In line with this, overall pupil size (across WM conditions) was larger in response to the more difficult AT Intervening task than the AS Intervening task. However, most of the patterns of pupil size we observed cannot be explained by task difficulty alone. For instance, pupil responses after the Intervening tasks were largest when participants had AT information in WM, despite more accurate recall in this condition than in the AS or VT WM conditions. A parsimonious explanation, consistent with the behavioral data, is that these pupil responses were more tightly linked to patterns of modality- and domain-based interference than the combined difficulty of the two tasks.

Interestingly, the initial pupil response during WM encoding was substantially larger in the AS WM condition than any of the other conditions. Participants did make relatively frequent behavioral errors in the AS WM condition, suggesting a task difficulty effect, but a similar performance difference between VT and VS recall (recall errors were more frequent in the VT condition) was not reflected in the pupil data. Since the pupil data did not simply mirror task performance, the larger pupil dilations in the AS WM condition suggest that encoding AS information in WM is especially effortful, perhaps due to the simultaneous recruitment of both complementary control networks. In other contexts, types of processing that require extra effort ultimately yield perceptual benefits. For instance, using context to fill in masked words supports speech understanding but requires effort (Winn & Moore, 2018), and integrating visual information in speech processing improves intelligibility (Sumby & Pollack, 1954) but may be effortful (Strand et al., 2020). Similarly, encoding AS information in WM appears to require

854 additional effort, but leveraging a visual-spatial framework to encode this information likely
855 results in the most robust memory storage.
856
857 *Insights from the neural data*
858 The fact that the Intervening tasks were always auditory limited the overall load on the
859 visual-biased control network in this study. In the visual WM conditions, auditory interference did
860 not rise to the level of affecting behavioral performance or effort requirements. In accord with the
861 strongest interference occurring when both tasks were auditory, the auditory Intervening tasks
862 elicited larger ERPs with visual as compared to auditory information held in WM. We posit that
863 this indicates greater resource availability for auditory perceptual processing when participants
864 were under a visual rather than auditory WM load. However, an alternative explanation could be
865 that stimuli presented during auditory WM encoding caused adaptation of neural responses,
866 resulting in suppressed amplitudes of auditory ERPs elicited by the onset of the Intervening task.
867 We do not suspect this was the case for two reasons. First, although the two-second gap between
868 the end of WM encoding and Intervening task onset was close enough that adaptation could
869 theoretically have had some effect on the ERPs (Budd et al., 1998), neural responses would have
870 likely been reset by the acoustic differences between the WM and Intervening task stimuli, as
871 observed in auditory oddball paradigms (Paavilainen, 2013). Second, adaptation cannot account
872 for differences in ERP amplitudes driven by differences in information domain (discussed below),
873 as the stimuli were physically identical across task domain conditions.
874 Intervening task ERPs were largest when the two tasks differed in both sensory modality
875 *and* information domain, even though similarity in information domain did not affect behavioral
876 performance or pupil size in the visual WM conditions. Importantly, the VT and VS WM tasks

877 used physically identical visual stimuli, and so differences in ERP amplitudes between these WM
878 conditions cannot be attributed to bottom-up stimulus salience. Instead, these results indicate that
879 the most neural resources were available for Intervening task processing when the two tasks
880 differed in both modality and domain, leading to the largest evoked responses. This provides neural
881 evidence that the WM control networks engaged by this task are organized not only according to
882 the sensory input channel, but also the domain of the information being represented.

883 Within the auditory WM conditions, no such pattern of domain-based interference was
884 found. Instead, ERPs elicited by both Intervening tasks tended to be slightly larger with AT than
885 AS information held in WM. One possible explanation relates to the particularly high degree of
886 effort required to encode AS information in WM, discussed in the previous section. If AS WM
887 encoding consumed a disproportionate amount of domain-general cognitive resources also needed
888 for attention in the Intervening task, ERP amplitudes in this WM condition may have been
889 generally suppressed.

890 The primary ERP component that was modulated by our task manipulations peaked around
891 200ms post-stimulus, and we have therefore referred to it as the "P2". However, since the
892 Intervening task stimuli were preceded by either visual or acoustically distinct auditory stimuli,
893 another possibility is that the large ERP positivity was actually the early phase of the novelty-P3
894 response, elicited by novel or unexpected sensory stimuli (Escera & Corral, 2007; Friedman et al.,
895 2001; Squires et al., 1975). The novelty-P3 can have a peak latency as early as 220ms and a scalp
896 distribution broadly consistent with the responses observed here (Escera et al., 1998). In this case,
897 ERP differences in the visual WM conditions could imply a novelty response to a change in the
898 information domain of the task. However, ERPs elicited by the second and third Intervening task
899 stimuli also had this weak N1/strong P2 morphology (see Supplemental Fig. 2), less consistent

900 with the novelty-P3. Alternatively, differences in auditory ERP morphology between the WM and
901 Intervening tasks may have resulted from acoustic differences between the stimuli (tone complexes
902 in the WM task, white noise bursts in the Intervening task). The data in Matsuo et al. (2019)
903 showed that P3 amplitudes were larger in ERPs elicited by white noise bursts than pure tones, and
904 parallels can be drawn between Intervening task ERPs in this study and the cortical responses to
905 noise-like fricative phonemes in Khalighinejad et al. (2017).

906 Alpha oscillatory power was modulated by the type of information being held in WM and
907 interrupted by the Intervening tasks. The largest alpha power changes from baseline were found at
908 central-parietal electrode sites. Alpha activity with this scalp distribution has been linked to
909 maintaining various types of information in WM, including visually presented letters and numbers
910 (Jensen et al., 2002; Klimesch et al., 1999; Schack & Klimesch, 2002), shapes (Herrmann et al.,
911 2004; Johnson et al., 2011), spatial information (Bastiaansen et al., 2002), and auditory stimuli
912 (Lim et al., 2015; Obleser et al., 2012; Vogt et al., 1998). In the auditory WM conditions, alpha
913 power was elevated during WM encoding and maintenance, but ongoing oscillations were
914 interrupted by the auditory Intervening tasks. This is consistent with previous work showing that
915 irrelevant visual stimuli interrupt alpha power during retention of both auditory and visual
916 information in WM (Hakim et al., 2020; Mishra et al., 2013). In the visual conditions, alpha power
917 was suppressed relative to baseline during WM encoding, likely because participants needed to
918 attend task-relevant visual stimuli (Foxe & Snyder, 2011; Jensen & Mazaheri, 2010). This alpha
919 suppression returned after participants performed the AS Intervening task, perhaps due to spatial
920 interference in the visual-biased network. The spatial interference from the AS Intervening task
921 suggested by the alpha and ERP data may be momentary in nature, as the AS Intervening task did
922 not require mapping stimulus positions into memory. Such ephemeral interference required the

temporal sensitivity of EEG to detect in this paradigm, but it may have meaningful processing consequences in real-world environments with many simultaneous sensory inputs and competing attentional and goal-oriented demands.

Conclusions

In a dual-task paradigm designed to rely on auditory- and visual-biased attention and WM networks, we found behavioral, autonomic, and electrophysiological signatures of interference when the two tasks drew upon shared neural control resources. Specifically, when perceptual and working memory tasks were matched in both sensory modality and information domain, behavioral performance was worst, pupil dilations were largest, and ERP amplitudes were suppressed—all indicating strong dual-task interference. These results support "resource-specific" theories of WM maintenance and indicate that WM control networks are segregated not only on the basis of the sensory information channel, but also based on whether the task requires spatial or temporal information processing. The bidirectionality of interference effects between the two tasks suggests that WM maintenance and perceptual processing relied on shared neural resources in this study, most likely including sensory-biased control networks in the lateral frontal cortex. Future neuroimaging studies should expand this paradigm to explore the dynamics of how LFC control networks support the interplay between WM maintenance and ongoing perceptual processing.

Acknowledgements

We would like to thank Jessica Tin and Yaminah Carter for their assistance with data collection and Matt Winn for discussions about interpreting the temporal dynamics of pupillometry data. The authors declare no competing interests. This work was supported by the Office of Naval Research

(Grant N000141812069) and the National Institute on Deafness and Other Communication Disorders (Grant T32-DC000038).

References

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings. Biological Sciences*, *273*(1592), 1339–1345. https://doi.org/10.1098/rspb.2005.3420

Arnell, K. M., & Jolicœur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 630–648. https://doi.org/10.1037/0096-1523.25.3.630

Bastiaansen, M. C. M., Posthuma, D., Groot, P. F. C., & de Geus, E. J. C. (2002). Event-related alpha and theta responses in a visuo-spatial working memory task. *Clinical Neurophysiology*, *113*(12), 1882–1893. https://doi.org/10.1016/S1388-2457(02)00303-6

Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener: Pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological Science*, *20*(11), 1313–1315. https://doi.org/10.1111/j.1467-9280.2009.02443.x

Braga, R. M., Hellyer, P. J., Wise, R. J. S., & Leech, R. (2017). Auditory and visual connectivity gradients in frontoparietal cortex. *Human Brain Mapping*, *38*(1), 255–270. https://doi.org/10.1002/hbm.23358

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology*, *31*(1), 51–68. https://doi.org/10.1016/S0167-8760(98)00040-3

Causse, M., Peysakhovich, V., & Fabre, E. F. (2016). High working memory load impairs language processing during a simulated piloting task: An ERP and pupillometry study. *Frontiers in Human Neuroscience*, *10*, 1–14. https://doi.org/10.3389/fnhum.2016.00240

Chun, M. M., & Potter, M. C. (2001). The attentional blink and task switching within and across modalities. In K. Shapiro (Ed.), *The limits of attention: Temporal constraints in human information processing*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198505150.001.0001/acprof-9780198505150

Cunillera, T., Fuentemilla, L., Periañez, J., Marco-Pallarès, J., Krämer, U. M., Càmara, E., Münte, T. F., & Rodríguez-Fornells, A. (2012). Brain oscillatory activity associated with task switching and feedback processing. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(1), 16–33. https://doi.org/10.3758/s13415-011-0075-5

Dynes, S. B. C., & Delgutte, B. (1992). Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. *Hearing Research*, *58*(1), 79–90. https://doi.org/10.1016/0378-5955(92)90011-B

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. https://doi.org/10.1162/089892998562997

Escera, C., & Corral, M. J. (2007). Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *Journal of Psychophysiology*, *21*(3–4), 251–264. https://doi.org/10.1027/0269-8803.21.34.251

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00154

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373. https://doi.org/10.1016/S0149-7634(01)00019-7

Gilzenrat, M. S., Nieuwenhuis, S., & Cohen, J. D. (2012). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function Mark. *Cogn Affect Behav Neurosci.*, *10*(2), 252–269. https://doi.org/10.3758/CABN.10.2.252.Pupil

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Hakim, N., Feldmann-Wüstefeld, T., Awh, E., & Vogel, E. K. (2020). Perturbing neural representations of working memory with task-irrelevant interruption. *Journal of Cognitive Neuroscience*, *32*(3), 558–569. https://doi.org/10.1162/jocn_a_01481

Herrmann, C. S., Senkowski, D., & Röttger, S. (2004). Phase-locking and amplitude modulations of EEG alpha: Two measures reflect different cognitive processes in a working memory task. *Experimental Psychology*, *51*(4), 311–318. https://doi.org/10.1027/1618-3169.51.4.311

Hsieh, S., & Allport, A. (1994). Shifting attention in a rapid visual search paradigm. *Perceptual and Motor Skills*, *79*(1), 315–335. https://doi.org/10.2466/pms.1994.79.1.315

Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex (New York, N.Y.: 1991)*, *12*(8), 877–882. https://doi.org/10.1093/cercor/12.8.877

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00186

Johnson, J. S., Sutterer, D. W., Acheson, D. J., Lewis-Peacock, J. A., & Postle, B. R. (2011). Increased alpha-band power during the retention of shapes and shape-location associations in visual short-term memory. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00128

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Khalighinejad, B., Cruzatto da Silva, G., & Mesgarani, N. (2017). Dynamic encoding of acoustic features in neural responses to continuous speech. *The Journal of Neuroscience*, *37*(8), 2176–2185. https://doi.org/10.1523/JNEUROSCI.2383-16.2017

Kim, S.-Y., Kim, M.-S., & Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *Proceedings of the National Academy of Sciences*, *102*(45), 16524–16529. https://doi.org/10.1073/pnas.0505454102

Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P., & Winkler, Th. (1999). `Paradoxical' alpha synchronization in a memory task. *Cognitive Brain Research*, *7*(4), 493–501. https://doi.org/10.1016/S0926-6410(98)00056-1

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research*, *323*, 81–90. https://doi.org/10.1016/j.heares.2015.02.004

Lim, S.-J., Carter, Y. D., Njoroge, J. M., Shinn-Cunningham, B. G., & Perrachione, T. K. (2021). Talker discontinuity disrupts attention to speech: Evidence from EEG and pupillometry. *Brain and Language*, *221*, 104996. https://doi.org/10.1016/j.bandl.2021.104996

Lim, S.-J., Wostmann, M., & Obleser, J. (2015). Selective attention to auditory memory neurally enhances perceptual precision. *Journal of Neuroscience*, *35*(49), 16094–16104. https://doi.org/10.1523/JNEUROSCI.2674-15.2015

Lin, C.-T., Chen, S.-A., Chiu, T.-T., Lin, H.-Z., & Ko, L.-W. (2011). Spatial and temporal EEG dynamics of dual-task driving performance. *Journal of NeuroEngineering and Rehabilitation*, *8*(11). https://doi.org/10.1186/1743-0003-8-11

López, M. E., Pusil, S., Pereda, E., Maestú, F., & Barceló, F. (2019). Dynamic low frequency EEG phase synchronization patterns during proactive control of task switching. *NeuroImage*, *186*, 70–82. https://doi.org/10.1016/j.neuroimage.2018.10.068

Marini, F., Scott, J., Aron, A. R., & Ester, E. F. (2017). Task-irrelevant distractors in the delay period interfere selectively with visual short-term memory for spatial locations. *Attention, Perception, & Psychophysics*, *79*(5), 1384–1392. https://doi.org/10.3758/s13414-017-1320-y

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Matsuo, M., Masuda, F., Sumi, Y., Takahashi, M., Yoshimura, A., Yamada, N., & Kadotani, H. (2019). Background music dependent reduction of aversive perception and its relation to P3 amplitude reduction and increased heart rate. *Frontiers in Human Neuroscience*, *13*, 184. https://doi.org/10.3389/fnhum.2019.00184

Mayer, A. R., Ryman, S. G., Hanlon, F. M., Dodd, A. B., & Ling, J. M. (2016). Look hear! The prefrontal cortex is stratified by modality of sensory input during multisensory cognitive control. *Cerebral Cortex*, bhw131. https://doi.org/10.1093/cercor/bhw131

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task: Physiological changes during sustained listening. *Psychophysiology*, *54*(2), 193–203. https://doi.org/10.1111/psyp.12772

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*(3), 211–253. https://doi.org/10.1006/cogp.2000.0736

Michalka, S. W., Kong, L., Rosen, M. L., Shinn-Cunningham, B. G., & Somers, D. C. (2015). Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. *Neuron*, *87*(4), 882–892. https://doi.org/10.1016/j.neuron.2015.07.028

Mishra, J., Zanto, T., Nilakantan, A., & Gazzaley, A. (2013). Comparable mechanisms of working memory interference by auditory and visual motion in youth and aging. *Neuropsychologia*, *51*(10), 1896–1906. https://doi.org/10.1016/j.neuropsychologia.2013.06.011

Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Current Eye Research*, *21*(1), 535–542. https://doi.org/10.1076/0271-3683(200007)2111-ZFT535

Morrison, N. M. V., Burnham, D., & Morrison, B. W. (2015). Cognitive load in cross-modal dual-task processing: Attention and cognitive load. *Applied Cognitive Psychology*, *29*(3), 436–444. https://doi.org/10.1002/acp.3122

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans: Indirect markers of locus coeruleus activity. *Psychophysiology*, *48*(11), 1532–1543. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*(3), 214–255. https://doi.org/10.1037/0033-295X.86.3.214

Nickerson, R. S. (Ed.). (1980). *Attention and performance VIII*. Erlbaum.

Noyce, A. L., Cestero, N., Michalka, S. W., Shinn-cunningham, B. G., & Somers, D. C. (2017). Sensory-biased and multiple-demand processing in human lateral frontal cortex. *Journal of Neuroscience*, *37*(36), 8755–8766. https://doi.org/10.1523/JNEUROSCI.0660-17.2017

Noyce, A. L., Lefco, R. W., Brissenden, J. A., Tobyne, S. M., Shinn-Cunningham, B. G., & Somers, D. C. (2022). Extended frontal networks for visual and auditory working memory. *Cerebral Cortex*, *32*(4), 855–869. https://doi.org/10.1093/cercor/bhab249

Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration☆. *Journal of Memory and Language*, *58*(3), 730–745. https://doi.org/10.1016/j.jml.2007.09.006

Obleser, J., Wostmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, *32*(36), 12376–12383. https://doi.org/10.1523/JNEUROSCI.4908-11.2012

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as

indicated by pupil dilation. *Hearing Research*, *351*, 68–79. https://doi.org/10.1016/j.heares.2017.05.012

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2011/156869

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, *88*(2), 109–123. https://doi.org/10.1016/j.ijpsycho.2013.03.015

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology*, *60*(2), 211–229. https://doi.org/10.1080/17470210600673818

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: Evidence from pupillometry and event-related fMRI. *NeuroImage*, *37*(3), 1017–1031. https://doi.org/10.1016/j.neuroimage.2007.04.066

Sauseng, P., Klimesch, W., Freunberger, R., Pecherstorfer, T., Hanslmayr, S., & Doppelmayr, M. (2006). Relevance of EEG alpha and theta oscillations during task switching. *Experimental Brain Research*, *170*(3), 295–301. https://doi.org/10.1007/s00221-005-0211-y

Scerra, V. E., & Brill, J. C. (2012). Effect of task modality on dual-task performance, response time, and ratings of operator workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1456–1460. https://doi.org/10.1177/1071181312561409

Schack, B., & Klimesch, W. (2002). Frequency characteristics of evoked and oscillatory electroencephalic activity in a human memory scanning task. *Neuroscience Letters*, *331*(2), 107–110. https://doi.org/10.1016/S0304-3940(02)00846-7

Serrien, D. J., Pogosyan, A. H., & Brown, P. (2004). Cortico-cortical coupling patterns during dual task performance. *Experimental Brain Research*, *157*, 79–84. https://doi.org/10.1007/s00221-003-1822-9

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*, *13*(11), 488–495. https://doi.org/10.1016/j.tics.2009.08.005

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

Stensaas, S. S., Eddington, D. K., & Dobelle, W. H. (1974). The topography and variability of the primary visual cortex in man. *Journal of Neurosurgery*, *40*(6), 747–755. https://doi.org/10.3171/jns.1974.40.6.0747

Strand, J. F., Brown, V. A., & Barbour, D. L. (2020). Talking points: A modulating circle increases listening effort without improving speech recognition in young adults. *Psychonomic Bulletin & Review*, *27*(3), 536–543. https://doi.org/10.3758/s13423-020-01713-y

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual topography of human intraparietal sulcus. *Journal of Neuroscience*, *27*(20), 5326–5337. https://doi.org/10.1523/JNEUROSCI.0991-07.2007

Tobyne, S. M., Osher, D. E., Michalka, S. W., & Somers, D. C. (2017). Sensory-biased attention networks in human lateral frontal cortex revealed by intrinsic functional connectivity. *NeuroImage*, *162*, 362–372. https://doi.org/10.1016/j.neuroimage.2017.08.020

van der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Kramer, J., Warmuth, E., Heekeren, H. R., & Wartenburger, I. (2010). Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology*, *47*(1), 158–169. https://doi.org/10.1111/j.1469-8986.2009.00884.x

Vogt, F., Klimesch, W., & Doppelmayr, M. (1998). High-frequency components in the alpha band and memory performance. *Journal of Clinical Neurophysiology*, *15*(2), 167–172. https://doi.org/10.1097/00004691-199803000-00011

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, *36*(4), 153–165. https://doi.org/10.1097/AUD.0000000000000145.The

Winn, M. B., & Moore, A. N. (2018). Pupillometry reveals that context benefit in speech perception can be disrupted by later-occurring sounds, especially in listeners with cochlear implants. *Trends in Hearing*, *22*, 2331216518808962. https://doi.org/10.1177/2331216518808962

Winn, M. B., & Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends in Hearing*, *25*, 233121652110276. https://doi.org/10.1177/23312165211027688

Zhao, S., Bury, G., Milne, A., & Chait, M. (2019). Pupillometry as an objective measure of sustained attention in young and older listeners. *Trends in Hearing*, *23*, 233121651988781. https://doi.org/10.1177/2331216519887815