Similarity in sensory modality and information domain impair processing in a dual-task context: Evidence from behavior, pupillometry, and EEG

Justin T. Fleming¹, J. Michelle Njoroge², Abigail L. Noyce³, Tyler K. Perrahione^{1,2}, and Barbara G. Shinn-Cunningham³

1. Speech and Hearing Bioscience and Technology Program, Harvard University, Cambridge, MA

2. Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA

3. Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA

Corresponding author:

Justin T. Fleming

164 Pillsbury Dr. SE

Minneapolis, MN 55455

jtf@umn.edu

Project Abstract

Perception and navigation frequently require us to maintain sensory information in memory, while also processing new sensory inputs as they arise. Recent fMRI research has uncovered regions in human frontal cortex that coordinate these diverse processes. Across various attention and working memory (WM) tasks, these regions can be separated into two distinct networks. Each shows a response preference for either auditory or visual stimuli, yet in addition, each can be flexibly recruited based on the information domain (i.e., temporal or spatial) of the task, regardless of sensory modality. Motivated by the sensory and functional specializations these networks display, we investigated whether dual-task interference is affected by similarity between the tasks in sensory modality (auditory or visual) and information domain. Participants performed a novel dual-task paradigm involving different combinations of WM and Intervening tasks, while two temporally sensitive physiological signals, pupillometry and electroencephalography (EEG), were measured. Convergent evidence from behavioral performance, pupil dilation amplitudes, and event-related potentials (ERPs) indicates that dual-task interference is greatest when the tasks match in both sensory modality and information domain. However, differences also arise in the patterns of dual-task interference across these metrics, highlighting the differential strengths and sensitivities of each. The results are consistent with increased interference when multiple tasks compete for shared cognitive control resources due to a common sensory modality or information domain.

Keywords: Dual-task, audio-visual, working memory, attention, pupillometry, EEG, alpha, lateral frontal cortex (LFC)

Introduction

Perceiving and acting in real-world environments requires a massive coordination of neural processes. We must extract spatial and temporal information about objects in our environment from multiple sensory modalities – particularly vision and audition – and store this information in working memory (WM) to guide future actions. Meanwhile, our perceptual systems must be flexible enough to process salient new sensory inputs as they arise without disrupting the information stored in WM. Recent fMRI research has identified sensory-biased networks in the human lateral frontal cortex (LFC) that may coordinate this interplay between sensory processing and WM. The LFC had previously been implicated in a wide array of cognitive functions, including working memory, sensorimotor control, and language processing, leading it to be frequently labelled a “multiple demand” structure (Assem et al., 2020; Blank et al., 2014; Duncan & Owen, 2000; Fedorenko et al., 2013; Postle et al., 2000). However, at the individual-participant level, distinct subregions within the LFC can be reliably identified with a preference for auditory or visual processing during attention and WM tasks (Braga et al., 2017; Mayer et al., 2016; Michalka et al., 2015; Noyce et al., 2017). Intrinsic functional connectivity analyses have revealed that these subregions are preferentially connected to the corresponding 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).

Importantly, processing in these LFC subregions is not strictly confined to information in the preferred sensory modality. Rather, the complementary LFC network can also be recruited depending on whether the information being processed is temporal or spatial in nature (Michalka et al., 2015). Vision and audition have complementary specializations for spatial and temporal processing, respectively. 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). At various stages along this pathway, visual neurons are sensitive to particular spatial features, such as edges, object orientation, spatial frequency, and texture (Hubel & Wiesel, 1959; Issa et al., 2000). Auditory spatial information, on the other hand, must be indirectly computed based on interaural cues. The peripheral auditory system is instead tuned to temporal information, with auditory nerve fibers capable of phase-locking to oscillating sound pressure waves (Dynes & Delgutte, 1992). This temporal sensitivity allows the auditory system to be organized around sound frequency (Da Costa et al., 2011; Merzenich & Brugge, 1973; Wessinger et al., 1997), and for perceptual judgement of inter-stimulus intervals to be much more precise in audition than vision (Lhamon & Goldstone, 1974). Reflecting this temporal specialization of the auditory system, the auditory-biased LFC network can be recruited by visual inputs if the task requires temporal information processing. Similarly, the visual-biased LFC network is recruited by auditory inputs if spatial information processing is required (Michalka et al., 2015). This flexible allocation of resources allows information to be processed by a network geared to the complementary strengths of each sensory modality.

Given what is known about recruitment in these sensory-biased cortical networks, in the present study we tested the prediction that dual-task interference will depend on whether the two tasks share the same sensory modality (auditory vs. visual) and information domain (temporal vs. spatial), thus causing the tasks to rely upon processing within the same cortical network. We developed a dual-task interference paradigm featuring WM and “Intervening” tasks. The WM task required participants to remember temporal or spatial information about a set of auditory or visual stimuli, while the Intervening task – presented during WM retention – required participants to make an immediate perceptual judgment about either the timing or spatial locations of auditory

stimuli. This design was inspired by “multiple resource theory,” which posits that two tasks will interfere with each other when they draw on the same pool of neural resources (Navon & Gopher, 1979; Nickerson, 1980; Wickens, 2002). 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.

To gain insights about the temporal dynamics of dual task interference in this paradigm, we combined behavioral measures with two temporally sensitive indexes of cognitive processing: pupillometry and electroencephalography (EEG). Task-evoked pupil dilations dynamically reflect task difficulty or the participant’s effort level during task performance (Causse et al., 2016; Gilzenrat et al., 2012; Murphy et al., 2011; Winn et al., 2015). EEG signals reveal different stages of processing through event-related potentials (ERPs) that are phase locked to sensory events (Luck, 2012), and more loosely coupled changes in oscillatory power in stereotypical frequencies that have been linked to different cognitive states (Ward, 2003). These techniques provide multiple viewpoints into dual-task interference with high temporal precision, but importantly, they lack the spatial resolution necessary to directly implicate the complementary sensory-biased cortical networks. Nonetheless, the WM tasks in this study were derived from the tasks Michalka et al. (2015) used to characterize the complementary LFC networks. Thus, we can say with relative confidence that the specific temporal and spatial WM tasks used in the current paradigm differentially engaged these networks.

In conditions with greater interference, we expected to observe the following phenomena: increased behavioral errors, both in WM retrieval and auditory perceptual judgments in the Intervening task; larger maximum pupil dilations; reduced ERP amplitudes, reflecting the strain

on neural resources needed to process the stimuli; and differing oscillatory signatures. In considering neural oscillations, we were specifically interested in the theta (4-8 Hz) and alpha (8-13 Hz) frequency bands, both of which have been implicated in task switching and WM maintenance (Bastiaansen et al., 2002; Cunillera et al., 2012; López et al., 2019; Obleser et al., 2012; Sauseng et al., 2006).

In brief, convergent evidence from behavior, pupillometry, and EEG indicated that interference was generally elevated when the WM and Intervening tasks relied on the same modality- and domain-biased attention and WM network. When the sensory modality differed between tasks, behavioral and autonomic signatures of interference were weak and indistinguishable based on WM domain. In these conditions, however, ERP amplitudes and alpha oscillations uncovered subtler patterns of domain-based interference. These results indicate that switching between complementary attention and WM networks comes at a lower cost than loading processing onto a single network.

Materials and Methods

Participants

Twenty-three healthy young adults completed all experimental procedures. Data from three participants was removed due to excessive noise in the pupillometry or EEG data, making for a final sample of twenty individuals (13 female, mean age 20.9 years, standard deviation 2.3 years). One additional participant was excluded from only the time-frequency analyses due to anomalous high-frequency noise in their EEG data. All participants had normal or corrected-to-normal visual acuity and no reported colorblindness. Participants who wore glasses for visual correction and

could not use contact lenses were excluded to avoid potential artifacts in the pupillometry data. All participants had clinically normal hearing, defined by tone detection thresholds below 20 dB HL at octave 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 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 $\pm 90^\circ$, $\pm 45^\circ$, 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 (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 was 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-second 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.

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.

Pilot testing with earlier versions of these WM tasks revealed necessary adjustments to prevent ceiling or floor effects in some conditions. For participants to perceive the different inter-stimulus intervals equally well, a larger 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). 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 identically 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 by 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. However, our analyses are primarily focused on comparing the four WM conditions *within* each Intervening task, as behavioral accuracy was unequal between the two tasks. By deemphasizing across-task comparisons, we sought to minimize the impact that potential differences in behavioral strategy could have 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 obfuscate

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, with the WM and (if present) Intervening task conditions held constant throughout the block. At the start of each block, an instruction screen indicated the sensory modality and relevant domain (temporal or spatial) of the upcoming trials. Participants were allowed to take untimed breaks between blocks. Participants performed one block of each WM and Intervening task combination 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, subject screening, and task practice. Participants practiced each variant of the WM and Intervening tasks in isolation until they understood the procedure, then 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. Session order 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/m²). 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/m². 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. WM task stimuli were only presented from the loudspeakers at ±90°, ±45°, and 0° azimuth. 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. These WM and Intervening task stimuli were chosen to be distinct, such that stimuli were clearly different between the two tasks. 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 the error rates on the WM and Intervening tasks. Reaction time estimates could not be made because 1) the position in the probe sequence at which a decision could be made differed based on whether the correct response was a “same” (final stimulus) or “different” (often earlier); and 2) within “different” trials, any of the second, third, or fourth stimuli could contain the change.

Statistical outcomes of behavioral performance were analyzed using logistic mixed effects regression models. For the WM task, the model included fixed effect terms for WM task modality, WM task domain, and Intervening task, and random effects terms to capture participant-specific intercepts for all three predictor variables. In Wilkinson notation, the model was specified as:

$$\begin{aligned} \text{logit}(\text{Error Rate}) \sim & \text{Modality}_{WM} * \text{Domain}_{WM} * \text{Int} \\ & + (1 + \text{Modality}_{WM} + \text{Domain}_{WM} + \text{Int} \mid ID) \end{aligned}$$

Due to the fundamental differences between the four WM conditions, differences in retrieval between these conditions were not of primary interest. Instead, we focused post-hoc analyses on how retrieval in each WM condition was impacted by modality and domain similarity with the Intervening task. To this end, post-hoc comparisons were only conducted within each WM task condition (e.g., examining differences in AT WM retrieval between the AT, AS, and No Intervening task conditions). Similarly, for Intervening task performance, we held the Intervening task fixed and compared error rates between the four WM task conditions. This analysis required us to compare across levels of both WM modality and WM domain; to facilitate this, the

Intervening task model was constructed with these factors collapsed into a single predictor variable, WM Condition:

$$\text{logit}(\text{Error Rate}) \sim \text{WM Condition} * \text{Int} + (1 + \text{WM Condition} + \text{Int} | \text{ID})$$

To investigate main effects and interactions at the group level, the coefficients from these models were supplied to a 3-way (WM task) or a 2-way (Intervening task) repeated measures ANOVA. Contrasts were initially treatment coded with base levels set as “auditory” for WM modality, “temporal” for WM domain, and “none” for Intervening task – these choices had no impact on the outcome of the group-level ANOVA. However, 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. All post-hoc tests were Bonferroni-Holm corrected for multiple comparisons.

Pupillometry data collection and analysis

Both pupils were tracked with a combination of the corneal reflection and a fit of the pupil border using SR Research algorithms. Prior to the start of each experimental session, eye position measurements were calibrated based on five fixation points (display center, $\pm 20^\circ$ azimuth on the horizontal plane, and $\pm 10^\circ$ 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, 2000 iterations of randomly shuffling the condition labels for each participant's data were performed; 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 actual 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 precluded whole-trial averaging); reject any epochs in which the data exceeded a 100 μ V 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 was 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 was 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 bandwidth) 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. 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 by 2000 random permutations of the condition labels.

Results

Intervening Task Performance

Performance accuracy on the Intervening tasks depended on both the Intervening task condition and the modality and domain of the information participants were holding in WM (Fig. 2). These effects were analyzed using a logistic mixed-effects model with fixed-effect terms of Intervening task (AT or AS) and WM task (AT, AS, VT, or VS). Random-effects terms included by-participant intercepts for both fixed effects. Group-level significance of the fixed effects and their interaction were tested in a Type-III ANOVA, with p-values based on the Satterthwaite approximation for degrees of freedom. This ANOVA revealed significant main effects of Intervening task ($\chi^2 = 32.7$, $df = 1$, $p = 1.07 \cdot 10^{-8}$) and WM task ($\chi^2 = 17.6$, $df = 3$, $p = 5.29 \cdot 10^{-4}$), as well as a significant interaction between these factors ($\chi^2 = 24.1$, $df = 3$, $p = 2.44 \cdot 10^{-5}$). The main effect of Intervening task indicates that performance was better overall on the AS Intervening task than the AT Intervening task. This difference simply reflects a limit on the difficulty of the AS Intervening task imposed by how close together the two loudspeakers could be placed. Since performance differed widely between the two Intervening tasks, we focused pairwise post-hoc tests on conditions in which the Intervening task was the same; differences between these conditions could only be caused by the type of information participants were holding in WM.

Figure 2: Intervening task error rates for each combination of WM task and Intervening task. Grey points represent individual participants, colored circles represent means, and error bars represent S.E.M. Chance performance is at an error trial proportion of 0.5. Asterisks indicate significant contrasts between WM task levels in the mixed-effects model, which were examined separately within each Intervening task. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ after Bonferroni-Holm correction.

As a means of post-hoc testing, we report the significance of contrasts in the mixed-effects model coding for each pair of WM tasks, with p-values adjusted for multiple comparisons using the Bonferroni-Holm correction. Participants made significantly more errors on the AT Intervening task 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, $p = 0.03$; Fig. 2, left). Accuracy on the AS Intervening task appeared to depend primarily on the sensory modality of the WM task, with participants making 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, $p = 0.021$; Fig. 2, right). Note that error rate floor effects were apparent on the AS Intervening task, which may have limited our ability to detect additional interference effects

based on WM domain. Nonetheless, participants generally committed more errors on the Intervening task when it matched the modality – and in the case of the AT Intervening task, the domain – of the information stored in WM.

Working Memory Task Performance

Participants were able to retrieve the information stored in WM at better-than-chance levels across all combinations of WM modality, WM domain, and Intervening task (Fig. 3). These factors (including conditions with no Intervening task) were included as fixed effects in a logistic mixed effects model of WM task retrieval errors. Random-effects terms included by-participant intercepts for all fixed effects, and group-level significance of the fixed effects was again tested using a Type-III ANOVA. This ANOVA revealed a significant interaction between WM task modality and domain ($\chi^2 = 206.4$, $df = 1$, $p < 10^{-15}$), reflecting overall lower error rates on the AT and VS WM tasks, in which the sensory modality was optimally suited for the information domain of the task (Fig. 3A). The ANOVA also revealed a three-way interaction between WM modality, WM domain, and Intervening task ($\chi^2 = 19.0$, $df = 2$, $p = 7.66 \cdot 10^{-5}$), indicating differential effects of the Intervening tasks depending on WM condition. These effects were explored via post-hoc testing, again restricted to conditions in which the task being measured – in this case, the WM task – was the same, with differences in WM retrieval now attributable only to the Intervening task type.

Figure 0: 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 WM task accuracy, Fig. 3B shows the difference between error rates when an AT or AS Intervening task was present and the corresponding condition with no Intervening task. When the WM task was AT, both of the auditory Intervening tasks significantly impaired WM retrieval ($p < 10^{-6}$ for both; top-left panel of Fig. 3B). 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 retaining AT information in WM relied solely on the auditory-temporal 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 were likely mapped into a representation in the visual-spatial WM network, protecting this information from auditory interference.

The fact that accuracy on the AS WM task was relatively poor might have impaired our ability to detect effects of the Intervening tasks in this condition, as retrieval was closer to chance before the Intervening task was added (Fig. 3A). To explore this possibility, we performed a median split of the AS WM task data in the no Intervening task condition, then tested for differences in the effects of the Intervening tasks between the better and worse performing participants. One might expect that the Intervening tasks would impair retrieval only among the better performers; this group started further from chance performance, resulting in more room for their performance to be affected by the Intervening tasks. However, this was not the case; neither Intervening task significantly altered AS WM task performance for either the better or worse performers, and the effects of the Intervening tasks did not differ between the two groups (uncorrected paired T-tests, all p-values > 0.27).

In some cases, visual WM retrieval was also modestly impaired by the auditory Intervening tasks. A small but 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. 3.6B). This might reflect domain-based WM interference with both tasks drawing on an auditory-temporal network, 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). This result was surprising given the lack

of modality or domain similarity between the WM and Intervening tasks. However, in this case the errors may reflect a task-switching cost as participants shifted from performing an auditory perceptual judgment to retrieving visual-spatial information from WM.

Pupil Dilations

Encoding information in WM and performing the auditory Intervening tasks both 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 (negative values on the time axis), peaked about 700 ms after the final WM encoding stimulus (time zero), and then gradually declined throughout WM retention (Fig. 4A, left panel). The onset of the AT and AS Intervening tasks (at 2 sec on the time axis) produced large secondary dilations, the amplitude of which was modulated by the information participants were holding in WM (Fig. 4A, middle and right panels).

Statistical differences between the pupil responses were assessed using non-parametric permutation testing, restricted to a time window spanning the WM retention phase (time zero and later). Tests was not conducted during WM task stimulus presentation because 1) the traces were baseline-corrected to zero at the start of this window, and 2) the interstimulus-intervals differed between auditory and visual WM conditions (see Methods), leading to different trace durations. In the absence of an Intervening task, AS WM encoding elicited a markedly larger pupil response than any of the other WM conditions (Fig. 3A). This may have reflected the generally higher difficulty of the AS WM task as compared to the AT WM task. However, while a similar performance difference was found between the VT and VS conditions, no difference in pupil dilations was detected between these conditions. Thus, these results may suggest an asymmetry

between the cognitive costs of encoding auditory information into a visual-spatial framework (AS WM) and encoding visual information into an auditory-temporal representation (VT WM).

Figure 4: Pupillometry results. **A**, Grand average z-scored pupil responses 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. 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.

Pupil dilations elicited by both auditory Intervening tasks were larger when the information stored in WM was auditory than when it was visual, mirroring the modality-driven interference effects observed in the behavioral data (Fig. 4A, middle and right panels). To more clearly separate the differential impact of the Intervening tasks on pupil diameter, we subtracted corresponding pupil traces with no Intervening task from each pupil trace with an Intervening task (Fig. 4B). As a preliminary analysis, we averaged these residual pupil traces across WM conditions and tested for a difference between the two Intervening tasks. This confirmed that traces elicited by the AT Intervening task were larger overall than those elicited by the AS Intervening task ($p < 0.05$ for all time points between 3.7 sec after WM retention onset and the end of the WM retention window; not shown), matching the higher behavioral error rates on the AT task.

Next, we compared the residual pupil traces between the four WM conditions within each Intervening task, similar to our treatment of the Intervening task behavioral data. Differential pupil dilations elicited by both Intervening tasks were larger when the information held in WM was auditory than when it was visual, as can also be observed in the raw traces in Fig. 4A. However, expressing the pupil dilations relative to conditions without an Intervening task revealed the largest differential response when the WM task was AT – significantly larger than when the WM task was AS – for both Intervening tasks. Further, the residual pupil traces were elevated prior to Intervening task onset in the AT WM condition, indicating larger pupil dilations during WM encoding when an Intervening task was upcoming. This suggests a preparatory modulation of task effort, with the AT WM information needing to be encoded with particularly high fidelity in order to survive interference from an auditory Intervening task. These results are broadly consistent with the behavioral data, in which the auditory Intervening tasks tended to be most costly when the WM task was AT. We expect that the AT WM condition maximally loaded auditory-biased attention

and WM networks, which were also necessary to process the stimuli in either of the auditory
Intervening tasks.

In the visual WM conditions, pupil dilations were insensitive to whether the WM task was
temporal or spatial. This result mirrors the behavioral data 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. Importantly, the patterns of pupil dilations we observed cannot
readily be explained by differences in WM task difficulty. Average pupil dilations were larger with
auditory than visual information held in WM, particularly when the WM task was AT. A pure
difficulty-based account would therefore predict that WM retrieval should be worst in the AT
condition, when in fact participants performed worst on the AS and VT WM tasks, in which the
sensory modality was sub-optimal for the information domain of the task. Instead, the pupil
dilations observed here likely reflect patterns of dual-task interference specific to the similarity in
modality and domain between the two tasks.

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 the
pattern of dual-task interference. One explanation for this could be that the AS Intervening task
did not recruit the visual-spatial attention network to the same extent as the AS WM task. The AS
Intervening task required only an immediate, relative judgement about sound locations, whereas
the AS WM task required participants to remember absolute sound positions by placing them in
allocentric reference frame. Thus, both Intervening tasks may have predominantly loaded on the
auditory-temporal attention network, resulting in similar behavioral and autonomic signatures of

interference. Nonetheless, the AT and AS Intervening tasks did require different types of auditory processing, 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 stimulus events in the WM and Intervening tasks. Grand average ERPs elicited by each stimulus event can be seen in Supplemental Fig. 1. ERPs elicited by stimuli in the encoding and probe sequences of the WM task were not systematically modulated by the Intervening tasks. However, ERPs elicited by the Intervening task stimuli – particularly the first sound in the Intervening task sequence – were modulated by the information participants were holding in WM. Unlike ERPs elicited by the auditory WM tasks, the Intervening task ERPs had a somewhat atypical morphology, including weak N1 and strong P2 components. This may have been due to the acoustic differences between the WM and Intervening task stimuli, or modulation of ERP morphology due to the fact that participants were holding information in WM. Regardless, both the modality and domain of the information in WM affected the amplitude of the P2 component of the Intervening task onset ERPs (Fig. 5, 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 P2 amplitudes were then modeled using a 3-way ANOVA with explanatory factors of WM modality, WM domain, and Intervening task. This analysis revealed a significant main effect of WM modality ($F = 7.27$, $df = 1$, $p = 0.014$) and a significant interaction between WM domain and Intervening task ($F = 10.83$, $df = 1$, $p = 0.004$), both of which will be examined in more detail below.

Figure 2: ERP results. **A**, Grand average ERPs elicited by the onset of 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. **C**, P2 peaks as a function of WM modality, averaged across WM domain and Intervening task. **D**, P2 peaks averaged across WM modality to illustrate the significant interaction between WM domain and Intervening task.

Starting with the main effect of WM modality, the amplitudes of P2 components elicited by both auditory Intervening tasks were larger when *visual* information was stored in WM than when the WM task was auditory (Fig. 5C). This could reflect a greater availability of auditory processing resources to respond to the Intervening task stimuli when these resources had not previously been taxed by an auditory WM task. An alternative explanation is that the auditory stimuli in the WM task caused habituation of the auditory Intervening task ERPs, resulting in suppressed P2 amplitudes. While the 2-second gap between WM encoding offset and Intervening task onset is close enough to reduce ERP amplitudes (Budd et al., 1998), such habituation would likely have been reset by the acoustic differences between the WM and Intervening task stimuli, similar to what occurs during auditory oddball paradigms (Paavilainen, 2013). In addition, ERPs elicited by the *final* AT Intervening task stimuli were modulated in a similar fashion to the onset ERPs (see Supplemental Fig. 1), which is inconsistent with a simple habituation account. A similar modulation of the final Intervening task ERPs was not observed in the AS task, but in this task, participants could sometimes make their decision before the final stimulus by comparing the first and second stimulus locations. Thus, the observed modulation of P2 amplitudes likely reflects resource availability for Intervening task processing, rather than auditory response habituation.

P2 amplitudes were also affected by domain similarity between the tasks, as evidenced by the significant interaction between WM domain and Intervening task. Specifically, P2s elicited by the first AS Intervening task stimulus were suppressed when the WM task was also spatial; this effect was, on average, smaller for the AT Intervening task (Fig. 5D). Although the three-way interaction involving WM task modality did not reach statistical significance, the two-way interaction appears to have been driven by the visual WM conditions, in which P2s were larger when the domain differed between the two tasks. To demonstrate this, we 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. The significant interaction between these terms remained in the visual WM conditions ($F = 10.77$, $df = 19$, $p = 0.004$; beneath a Bonferroni-Holm adjusted alpha criterion to account for the two additional ANOVAs, assuming an initial alpha criterion of $p = 0.05$). In the auditory WM conditions, no significant main effects or interactions were detected (even when uncorrected for multiple comparisons). In sum, P2 components of the ERPs scaled inversely with dual-task similarity in modality and domain. P2s amplitudes were generally higher when the tasks were mismatched on either dimension, leading to the largest P2s when the tasks differed in both modality and domain.

Alpha-band oscillatory activity

During WM encoding, oscillatory power in the alpha (8-13 Hz) frequency band steadily increased relative to baseline in the auditory WM conditions, and decreased relative to baseline in the visual WM conditions (Supplemental Fig. 2). These patterns of alpha power persisted into the WM retention window, shown in Fig. 6. Preliminary analysis of alpha power in the retention window using cluster-based permutation testing revealed no significant differences based on the domain of the WM task, so data were collapsed across this dimension for the following analyses. The grand average power spectrum during the pre-trial baseline, showing a clear peak in the alpha band at a parietal electrode site, can be found in Supplemental Fig. 3.

Figure 6: Alpha power results. 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 and for auditory (A) and visual (C) WM. 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. A timing error in the experiment code shortened the memory retention window for two subjects by approximately 100 ms, and so all time-frequency responses were additionally truncated to match this shortened duration. B, Grand average alpha power time courses (left panels) and statistical comparisons (right panels) for the auditory WM conditions. D, Same as B, but for the visual 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 for the comparisons shown in the topoplots to the right. When significant clusters were detected, the alpha power difference was plotted only in the significant time region; otherwise, the topoplot time regions span the entire cluster test window. * indicates channel membership in a significant cluster ($p < 0.05$).

In the auditory WM tasks, alpha power remained elevated relative to baseline throughout the memory retention window when no Intervening task was present (Fig. 6A, top panel). However, the onset of both the AT and AS Intervening tasks suppressed these ongoing alpha oscillations (middle and bottom panels). We examined this effect by conducting cluster-based permutation tests on the alpha power time courses (see Materials and Methods) in two time regions of interest: one spanning the onset of the Intervening task through the end of the Intervening task response window (2 to 3.5 seconds into WM retention), and another from 3.5 to 5 sec, during which participants needed to start retrieving the information encoded in WM in anticipation of the probe sequence. During both of the Intervening tasks (in the earlier window), alpha power was significantly reduced relative to the no Intervening task condition (AT Intervening task, $p = 0.005$; AS Intervening task, $p = 0.003$; Fig 6B). With the statistical threshold for generating the initial clusters set at $p = 0.01$, no differences were found between alpha power in the two Intervening tasks. However, relaxing this threshold to $p = 0.05$ revealed a late AT vs. AS Intervening task cluster between 3.14 and 3.45 sec ($p = 0.026$; not shown). This indicates that the alpha power reduction lasted longer when the Intervening task was AT than when it was AS, probably because participants could make their perceptual judgment earlier in the AS Intervening task than the AT Intervening task. In the auditory WM conditions, no significant clusters were found in the later (3.5-5 sec) time window.

In the visual WM conditions without an Intervening task, alpha power gradually returned from the suppression during WM encoding to near-baseline levels during WM retention (Fig. 6C, top panel). On trials with an AS Intervening task, there was a tendency for alpha to become

suppressed again after the conclusion of the Intervening task (bottom panel). We examined these alpha power time courses using cluster-based permutation testing restricted to the same time windows as in the auditory WM conditions: 2 to 3.5 sec and 3.5 to 5 sec from the start of the WM retention phase. Unlike the auditory conditions, no significant differences were observed during the Intervening tasks. However, in the later window, when the Intervening task was AS, alpha power was significantly reduced relative to when there was no Intervening task ($p = 0.011$; Fig. 6D). This alpha reduction did not reach significance following the AT Intervening task. When the cluster formation threshold was again relaxed to $p < 0.05$, alpha power was found to be significantly lower following the AS Intervening task than the AT task ($p = 0.027$, not shown). In other words, the alpha suppression present during visual WM encoding resumed following the spatial Intervening task, possibly representing a signature of domain-driven interference.

Finally, the presence of either Intervening task also caused an increase in frontal theta (4-7 Hz) power throughout the memory retention window (Supplemental Fig. 4). 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 was trending 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. Frontal-midline theta power and theta-band phase coupling in fronto-parietal networks have been linked to task-switching (Cunillera et al., 2012; López et al., 2019; Sauseng et al., 2006) and high levels of cognitive load (Gevins, 1997; McEvoy et al., 2000), both of which were likely relevant in conditions with an Intervening task.

Discussion

We used a dual-task paradigm, comprising WM and auditory Intervening tasks, to investigate processing interference as a function of whether the two tasks matched in sensory modality and information domain. Convergent evidence from behavior and pupillometry showed that processing was negatively impacted when the two tasks drew upon shared processing resources. Specifically, this manifested as the highest Intervening task error rates, poorest memory retrieval, and largest pupil dilations when both tasks were auditory-temporal. We expected auditory-temporal WM information to load exclusively on auditory-biased attention and WM networks, likely increasing interference with the auditory Intervening tasks. In contrast, auditory-spatial and visual WM information was likely mapped to and maintained in visual-spatial networks, protecting it from interference by the auditory Intervening tasks. Measures of interference were smallest in the visual WM conditions, and they did not differ depending on whether the WM task was temporal or spatial. This suggests that interference was minimal when competing tasks were presented in different sensory modalities.

In these conditions with reduced competition for shared neural resources, subtler patterns of dual-task interference were nonetheless revealed in the neural data. The amplitudes of the P2 components of auditory ERPs elicited by Intervening task onset tended to be suppressed by dual-task similarity in either modality or domain. The largest P2 components were observed in visual WM conditions when the information domain was mismatched between the two tasks. Ongoing alpha (8-13 Hz) oscillations during WM maintenance were also modulated by the presence of an Intervening task. In the auditory WM conditions, alpha power was strongly interrupted by both the AT and the AS Intervening tasks. In the visual WM conditions, alpha power was suppressed following selectively the auditory-spatial Intervening task, possibly indicating the need to re-

engage WM mechanisms more strongly after interruption of visual WM rehearsal by spatial processing in a different sensory modality.

Similarity in sensory modality leads to increased dual-task interference

While this is the first study we are aware of to investigate how both sensory modality and information domain influence dual-task interference, several previous behavioral studies have determined that interference depends on some axis of similarity between two simultaneous tasks. For instance, Kim et al. (2005) reported increased Stroop task interference when WM was loaded with verbal information that overlapped with the Stroop target. On the other hand, there was no effect of loading WM with spatial information irrelevant to the Stroop task, and loading WM with information similar to the Stroop task distractor actually reduced interference. Similarly, when participants have to retrieve both the orientation and location of visual items, visual distractors that differ from the target only in orientation selectively impair retrieval of target orientation, but not location (Marini et al., 2017). In the realm of phonological WM, distractor words presented during WM retention selectively interfere with recall of words with which they share phonological properties (Oberauer & Lange, 2008).

In line with the current study, previous work has specifically implicated similarity in sensory modality as a determinant of distractor and dual-task interference. For instance, auditory and visual discrimination thresholds (for pitch and contrast, respectively) are unaffected by concurrent, irrelevant stimuli in the opposite modality, but are considerably higher in the presence of concurrent stimuli in the same modality (Alais et al., 2006). In dual-task paradigms, increased interference has been observed when both tasks are tactile, auditory, or visual, as compared to conditions in which the tasks were presented in different modalities (Morrison et al., 2015; Scerra

& Brill, 2012). These results lend support to multiple resource theory, which states that interference between two tasks depends on whether they compete for the same set of limited neural resources.

Dual-task interference: A single central or task-specific bottleneck?

The neural underpinnings of dual-task interference have been the topic of substantial debate. An opposing view to the multiple resource theory posits that a general, “central bottleneck” limits performance independently of the specific processing demands of the component tasks (Pashler, 1998); a number of neuroimaging studies have provided support for this view (Dux et al., 2006; Marois & Ivanoff, 2005; Tombu et al., 2011). However, contrasting studies more in line with multiple resource theory have shown that interference is marked by increased activity in the brain regions involved in processing the component tasks, without recruitment of any additional areas in dual-task conditions (Adcock et al., 2000; Bunge et al., 2000). This has also been observed at the level of single neurons, with dual-task interference being linked to reduced representational fidelity in a shared population of lateral prefrontal cortex neurons required for both tasks (Watanabe & Funahashi, 2014).

In the current study, we attempted to differentially load auditory- and visual-biased cortical networks with our dual-task paradigm. Recruitment of these two networks for attention and WM processing is clearly differentiable based on both sensory modality and information domain (Michalka et al., 2015, 2016; Noyce et al., 2017), supporting the view that they are physiologically distinct neural resources. On the other hand, each network is recruited by a variety of tasks within its preferred modality and domain. Thus, two related but unidentical tasks could interfere with each other if they both load onto the same cortical network, consistent with a “central

bottleneck.” The coexistence of these properties suggests that the central bottleneck and multiple resource accounts of dual-task interference may not be entirely dichotomous, an idea supported by the present results. Interference patterns measured with behavior and pupillometry were generally similar between the AT and AS Intervening tasks, indicating that both Intervening tasks loaded predominantly onto the same network (likely the auditory-temporal network), despite one involving temporal processing and the other spatial processing (i.e., an apparent central bottleneck). In the WM task on the other hand, dual-task interference patterns were strikingly different between the AT and AS WM conditions. The AS WM task required participants to map absolute sound locations into exocentric coordinates, which likely drove the visual-spatial network more than the AS Intervening task, which only required an immediate comparison of two sets of interaural cues. In other words, the AS WM task was likely able to recruit a separate pool of neural resources, reducing interference from the auditory Intervening tasks; this result is more consistent with multiple resource theory. Thus, there may not be a clear and generalizable resolution between these two theories; rather, dual-task interference effects likely depend on relationships between the specific tasks used, and the specific brain regions and networks these tasks recruit.

Another noteworthy aspect of the results presented here is that the processing costs of loading onto a single network (measured as increased behavioral interference, larger pupil diameter, and ERP amplitude suppression) were greater than the costs of switching between networks. This outcome is not entirely obvious, as task-switching also incurs a processing cost (Arnell & Jolicoeur, 1999; Chun & Potter, 2001; Hsieh & Allport, 1994; Meiran et al., 2000). Our auditory Intervening tasks did modestly impair visual WM retrieval, consistent with a behavioral cost of switching between the complementary networks. However, this effect was relatively weak compared to the behavioral (and physiological) costs when the tasks loaded onto shared network

resources. Thus, the ability to simultaneously leverage complementary attention and WM networks appears to confer a processing benefit, which may be especially relevant in complex, multisensory, real-world environments.

Separate contributions of behavioral, pupillometry, and EEG measures

Measures of behavior, pupillometry, and EEG in this study each offered unique insights into the dual-task interference at play. Pupillometry was included because it can serve as a sensitive biomarker 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 when the signal-to-noise ratio is high enough for both groups to achieve ceiling performance. This is thought to reflect a heightened degree of listening effort, not captured by behavioral measures alone (Ohlenforst et al., 2017). Pupil dilations have also been shown to correlate with participants' subjective assessments of task difficulty (Koelewijn et al., 2015). In the present study, this benefit manifested in our ability to detect differences between the AT and AS WM conditions when the Intervening task was AS. This Intervening task was easy enough that a majority of participants did not make errors on it in any WM condition. However, substantial differences in the amplitude of pupil dilations elicited by the AS Intervening task indicated that the information in WM did cause processing interference, which participants overcame to achieve excellent performance.

In addition, behavioral performance on each trial is often a single-sample snapshot, encompassing all the processing dynamics present in the task. Pupillometry allows for more fine-grained dissection of the time course of these dynamics. For instance, pupil data in this study revealed that participants were modulating their effort level to encode AT information in WM

when an auditory Intervening task would be present, indicating preparation for a challenging interference condition. Analysis of the temporal dynamics of pupil responses is becoming more common in other fields as well, including the study of cognitive processes that unfold through time like decision making (Satterthwaite et al., 2007) and the perception of temporally complex inputs like speech and music (Jagiello et al., 2019; Winn & Moore, 2018). Conversely, though, the behavioral data provided insights that would have been difficult to glean from pupillometry alone. Pupil responses elicited by the Intervening task stimuli were drastically different from the no Intervening task conditions, making it difficult to interpret the magnitude of the interference effects. The behavioral data puts this in more concrete terms; we could measure exactly how much performance on each task was influenced by the other. Because of these complementary benefits, we argue that pupillometry data are best interpreted alongside corresponding behavioral measures.

In the visual WM conditions, interference by the auditory Intervening tasks was generally low, and neither behavioral nor pupil responses differed based on whether the information held in WM was temporal or spatial. This finding was counter to our expectations; for instance, we expected to observe mild interference when the Intervening and WM tasks were both spatial. Auditory and visual spatial processing are known to be represented in the EEG response, often in the form of lateralized alpha oscillations at parietal scalp locations (Bednar & Lalor, 2018; Deng et al., 2020; Kelly et al., 2006; Klatt et al., 2018; Thut et al., 2006; Tune et al., 2018; Worden et al., 2000). Given the sensitivity of EEG to spatial processing, we expected to observe domain-based interference patterns in our EEG data when the WM and Intervening tasks were both spatial. Indeed, ERPs and alpha modulations proved particularly sensitive to these subtle interference patterns, which did not affect error rates or pupil responses. The P2 components of ERPs were largest when the tasks were presented in different modalities and, in the case of the visual WM

conditions, when the information domains of the tasks were also mismatched. This pattern is consistent with greater availability of processing resources when the tasks did not compete for shared modality- and domain-biased neural resources (Kok, 1997).

We found that alpha power was modulated by the presence of an Intervening task and by its information domain. This alpha had a more anterior and centralized distribution than the lateralized alpha often implicated in spatial attention (Fig. 6). Alpha power with the scalp distribution we found in this study has been linked to various types of WM load, including visually presented letters and numbers (Jensen et al., 2002; Klimesch et al., 1999; Schack & Klimesch, 2002), shape information (Herrmann et al., 2004; Johnson et al., 2011), spatial information (Bastiaansen et al., 2002), and auditory stimuli (Lim et al., 2015; Obleser et al., 2012; Vogt et al., 1998). In the auditory WM conditions, the auditory Intervening tasks interrupted this alpha activity during WM retention. This is consistent with previous work showing that distracting visual stimuli interrupt alpha power during retention of both auditory and visual information in WM (Hakim et al., 2020; Mishra et al., 2013). In the visual WM conditions, alpha power was suppressed relative to baseline during WM encoding, perhaps due to the need to attend task-relevant visual stimuli (Foxe & Snyder, 2011; Jensen & Mazaheri, 2010). In these conditions, there was no significant change in alpha power during Intervening task presentation; however, following the spatial Intervening task, the alpha suppression resumed. The functional implications of these alpha suppressive effects, and why their pattern differed between the auditory and visual WM conditions, are topics warranting targeted future investigation.

Conclusions and future directions

In a dual-task paradigm designed to drive auditory- and visual-biased attention and WM networks, we reported behavioral, autonomic, and electrophysiological signatures of interference when the tasks drew upon shared neural control resources. Future studies should implement a similar paradigm with fMRI to pinpoint the neural networks involved in this kind of dual-task interference, as well as invert the paradigm used here – using visual rather than auditory Intervening tasks – to test whether interference patterns flip under opposite network demands.

Acknowledgements

We would like to thank Jessica Tin for her assistance with data collection and Dr. Matt Winn for discussions about the 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

- Adcock, R. A., Constable, R. T., Gore, J. C., & Goldman-Rakic, P. S. (2000). Functional neuroanatomy of executive processes involved in dual-task performance. *Proceedings of the National Academy of Sciences*, 97(7), 3567–3572. <https://doi.org/10.1073/pnas.97.7.3567>
- 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>
- Assem, M., Glasser, M. F., Van Essen, D. C., & Duncan, J. (2020). A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cerebral Cortex*, 30(8), 4361–4380. <https://doi.org/10.1093/cercor/bhaa023>
- 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](https://doi.org/10.1016/S1388-2457(02)00303-6)
- Bednar, A., & Lalor, E. C. (2018). Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. *NeuroImage*, 181, 683–691. <https://doi.org/10.1016/j.neuroimage.2018.07.054>
- Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, 112(5), 1105–1118. <https://doi.org/10.1152/jn.00884.2013>
- 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>

- 906 Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory
907 event-related potential with stimulus repetition: Habituation vs. refractoriness. *International*
908 *Journal of Psychophysiology*, 31(1), 51–68. [https://doi.org/10.1016/S0167-8760\(98\)00040-3](https://doi.org/10.1016/S0167-8760(98)00040-3)
- 909 Bunge, S. A., Klingberg, T., Jacobsen, R. B., & Gabrieli, J. D. E. (2000). A resource model of the neural
910 basis of executive working memory. *Proceedings of the National Academy of Sciences*, 97(7),
911 3573–3578. <https://doi.org/10.1073/pnas.97.7.3573>
- 912 Causse, M., Peysakhovich, V., & Fabre, E. F. (2016). High Working Memory Load Impairs Language
913 Processing during a Simulated Piloting Task: An ERP and Pupillometry Study. *Frontiers in*
914 *Human Neuroscience*, 10, 1–14. <https://doi.org/10.3389/fnhum.2016.00240>
- 915 Chun, M. M., & Potter, M. C. (2001). The attentional blink and task switching within and across
916 modalities. In K. Shapiro (Ed.), *The limits of attention: Temporal constraints in human*
917 *information processing*. Oxford University Press.
918 [http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198505150.001.0001/acprof-](http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198505150.001.0001/acprof-9780198505150)
919 [9780198505150](http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198505150.001.0001/acprof-9780198505150)
- 920 Cunillera, T., Fuentemilla, L., Periañez, J., Marco-Pallarès, J., Krämer, U. M., Càmarà, E., Münte, T. F.,
921 & Rodríguez-Fornells, A. (2012). Brain oscillatory activity associated with task switching and
922 feedback processing. *Cognitive, Affective, & Behavioral Neuroscience*, 12(1), 16–33.
923 <https://doi.org/10.3758/s13415-011-0075-5>
- 924 Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011).
925 Human primary auditory cortex follows the shape of Heschl’s gyrus. *Journal of Neuroscience*,
926 31(40), 14067–14075. <https://doi.org/10.1523/JNEUROSCI.2000-11.2011>
- 927 Deng, Y., Choi, I., & Shinn-Cunningham, B. (2020). Topographic specificity of alpha power during
928 auditory spatial attention. *NeuroImage*, 207, 116360.
929 <https://doi.org/10.1016/j.neuroimage.2019.116360>
- 930 Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal loberecruited by diverse
931 cognitive demands. *Trends in Neurosciences*, 23(10), 475–483.

932 Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a Central Bottleneck of
933 Information Processing with Time-Resolved fMRI. *Neuron*, 52(6), 1109–1120.
934 <https://doi.org/10.1016/j.neuron.2006.11.009>

935 Dynes, S. B. C., & Delgutte, B. (1992). Phase-locking of auditory-nerve discharges to sinusoidal electric
936 stimulation of the cochlea. *Hearing Research*, 58(1), 79–90. [https://doi.org/10.1016/0378-](https://doi.org/10.1016/0378-5955(92)90011-B)
937 [5955\(92\)90011-B](https://doi.org/10.1016/0378-5955(92)90011-B)

938 Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal
939 and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of*
940 *America*, 110(41), 16616–16621. <https://doi.org/10.1073/pnas.1315235110>

941 Foxe, J. J., & Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression
942 Mechanism during Selective Attention. *Frontiers in Psychology*, 2.
943 <https://doi.org/10.3389/fpsyg.2011.00154>

944 Gevins, A. (1997). High-resolution EEG mapping of cortical activation related to working memory:
945 Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, 7(4), 374–385.
946 <https://doi.org/10.1093/cercor/7.4.374>

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

950 Hakim, N., Feldmann-Wüstefeld, T., Awh, E., & Vogel, E. K. (2020). Perturbing Neural Representations
951 of Working Memory with Task-irrelevant Interruption. *Journal of Cognitive Neuroscience*, 32(3),
952 558–569. https://doi.org/10.1162/jocn_a_01481

953 Herrmann, C. S., Senkowski, D., & Röttger, S. (2004). Phase-Locking and Amplitude Modulations of
954 EEG Alpha: Two Measures Reflect Different Cognitive Processes in a Working Memory Task.
955 *Experimental Psychology*, 51(4), 311–318. <https://doi.org/10.1027/1618-3169.51.4.311>

956 Hsieh, S., & Allport, A. (1994). Shifting Attention in a Rapid Visual Search Paradigm. *Perceptual and*
957 *Motor Skills*, 79(1), 315–335. <https://doi.org/10.2466/pms.1994.79.1.315>

958 Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The*
959 *Journal of Physiology*, 148(3), 574–591. <https://doi.org/10.1113/jphysiol.1959.sp006308>

960 Issa, N. P., Trepel, C., & Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *The Journal*
961 *of Neuroscience*, 20(22), 8504–8514. <https://doi.org/10.1523/JNEUROSCI.20-22-08504.2000>

962 Jagiello, R., Pomper, U., Yoneya, M., Zhao, S., & Chait, M. (2019). Rapid Brain Responses to Familiar
963 vs. Unfamiliar Music – an EEG and Pupillometry study. *Scientific Reports*, 9(1), 15570.
964 <https://doi.org/10.1038/s41598-019-51759-9>

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

968 Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity:
969 Gating by Inhibition. *Frontiers in Human Neuroscience*, 4.
970 <https://doi.org/10.3389/fnhum.2010.00186>

971 Johnson, J. S., Sutterer, D. W., Acheson, D. J., Lewis-Peacock, J. A., & Postle, B. R. (2011). Increased
972 Alpha-Band Power during the Retention of Shapes and Shape-Location Associations in Visual
973 Short-Term Memory. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00128>

974 Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in Alpha Oscillatory Power Reflect
975 an Active Retinotopic Mechanism for Distracter Suppression During Sustained Visuospatial
976 Attention. *Journal of Neurophysiology*, 95(6), 3844–3851. <https://doi.org/10.1152/jn.01234.2005>

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

980 Klatt, L.-I., Getzmann, S., Wascher, E., & Schneider, D. (2018). The contribution of selective spatial
981 attention to sound detection and sound localization: Evidence from event-related potentials and
982 lateralized alpha oscillations. *Biological Psychology*, 138, 133–145.
983 <https://doi.org/10.1016/j.biopsycho.2018.08.019>

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

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

990 Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: A review and synthesis.
991 *Biological Psychology*, 45(1–3), 19–56. [https://doi.org/10.1016/s0301-0511\(96\)05221-0](https://doi.org/10.1016/s0301-0511(96)05221-0)

992 Lhamon, W. T., & Goldstone, S. (1974). Studies of auditory-visual differences in human time judgment
993 2: More transmitted information with sounds than lights. *Perceptual and Motor Skills*, 39(1),
994 295–307. <https://doi.org/10.2466/pms.1974.39.1.295>

995 Lim, S.-J., Wostmann, M., & Obleser, J. (2015). Selective Attention to Auditory Memory Neurally
996 Enhances Perceptual Precision. *Journal of Neuroscience*, 35(49), 16094–16104.
997 <https://doi.org/10.1523/JNEUROSCI.2674-15.2015>

998 López, M. E., Pusil, S., Pereda, E., Maestú, F., & Barceló, F. (2019). Dynamic low frequency EEG phase
999 synchronization patterns during proactive control of task switching. *NeuroImage*, 186, 70–82.
1000 <https://doi.org/10.1016/j.neuroimage.2018.10.068>

1001 Luck, S. J. (2012). Event-related potentials. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D.
1002 Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 1:*
1003 *Foundations, planning, measures, and psychometrics*. (pp. 523–546). American Psychological
1004 Association. <https://doi.org/10.1037/13619-028>

1005 Marini, F., Scott, J., Aron, A. R., & Ester, E. F. (2017). Task-irrelevant distractors in the delay period
1006 interfere selectively with visual short-term memory for spatial locations. *Attention, Perception, &*
1007 *Psychophysics*, 79(5), 1384–1392. <https://doi.org/10.3758/s13414-017-1320-y>

1008 Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in*
1009 *Cognitive Sciences*, 9(6), 296–305. <https://doi.org/10.1016/j.tics.2005.04.010>

1010 Mayer, A. R., Ryman, S. G., Hanlon, F. M., Dodd, A. B., & Ling, J. M. (2016). Look Hear! The
 1011 Prefrontal Cortex is Stratified by Modality of Sensory Input During Multisensory Cognitive
 1012 Control. *Cerebral Cortex*, bhw131. <https://doi.org/10.1093/cercor/bhw131>

1013 McEvoy, L. K., Smith, M. E., & Gevins, A. (2000). Test–retest reliability of cognitive EEG. *Clinical*
 1014 *Neurophysiology*, 111(3), 457–463. [https://doi.org/10.1016/S1388-2457\(99\)00258-8](https://doi.org/10.1016/S1388-2457(99)00258-8)

1015 Meiran, N., Chorev, Z., & Sapir, A. (2000). Component Processes in Task Switching. *Cognitive*
 1016 *Psychology*, 41(3), 211–253. <https://doi.org/10.1006/cogp.2000.0736>

1017 Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition on the superior
 1018 temporal plane of the macaque monkey. *Brain Research*, 50(2), 275–296.
 1019 [https://doi.org/10.1016/0006-8993\(73\)90731-2](https://doi.org/10.1016/0006-8993(73)90731-2)

1020 Michalka, S. W., Kong, L., Rosen, M. L., Shinn-Cunningham, B. G., & Somers, D. C. (2015). Short-
 1021 Term Memory for Space and Time Flexibly Recruit Complementary Sensory-Biased Frontal
 1022 Lobe Attention Networks. *Neuron*, 87(4), 882–892. <https://doi.org/10.1016/j.neuron.2015.07.028>

1023 Michalka, S. W., Rosen, M. L., Kong, L., Shinn-Cunningham, B. G., & Somers, D. C. (2016). Auditory
 1024 Spatial Coding Flexibly Recruits Anterior, but Not Posterior, Visuotopic Parietal Cortex.
 1025 *Cerebral Cortex (New York, N.Y.: 1991)*, 26(3), 1302–1308.
 1026 <https://doi.org/10.1093/cercor/bhv303>

1027 Mishra, J., Zanto, T., Nilakantan, A., & Gazzaley, A. (2013). Comparable mechanisms of working
 1028 memory interference by auditory and visual motion in youth and aging. *Neuropsychologia*,
 1029 51(10), 1896–1906. <https://doi.org/10.1016/j.neuropsychologia.2013.06.011>

1030 Morrison, N. M. V., Burnham, D., & Morrison, B. W. (2015). Cognitive Load in Cross-Modal Dual-Task
 1031 Processing: Attention and cognitive load. *Applied Cognitive Psychology*, 29(3), 436–444.
 1032 <https://doi.org/10.1002/acp.3122>

1033 Murphy, P. R., Robertson, I. H., Balsters, J. H., & O’connell, R. G. (2011). Pupillometry and P3 index the
 1034 locus coeruleus-noradrenergic arousal function in humans: Indirect markers of locus coeruleus

1035 activity. *Psychophysiology*, 48(11), 1532–1543. <https://doi.org/10.1111/j.1469->
1036 8986.2011.01226.x

1037 Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological*
1038 *Review*, 86(3), 214–255. <https://doi.org/10.1037/0033-295X.86.3.214>

1039 Nickerson, R. S. (Ed.). (1980). *Attention and performance VIII: Proceedings of the Eighth International*
1040 *Symposium on Attention and Performance, Princeton, New Jersey, USA, August 20-25, 1978.*
1041 Erlbaum.

1042 Noyce, A. L., Cestero, N., Michalka, S. W., Shinn-cunningham, B. G., & Somers, D. C. (2017). *Sensory-*
1043 *Biased and Multiple-Demand Processing in Human Lateral Frontal Cortex.* 37(36), 8755–8766.
1044 <https://doi.org/10.1523/JNEUROSCI.0660-17.2017>

1045 Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-
1046 based confusion, feature overwriting, and feature migration☆. *Journal of Memory and Language*,
1047 58(3), 730–745. <https://doi.org/10.1016/j.jml.2007.09.006>

1048 Obleser, J., Wostmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse Listening Conditions
1049 and Memory Load Drive a Common Alpha Oscillatory Network. *Journal of Neuroscience*,
1050 32(36), 12376–12383. <https://doi.org/10.1523/JNEUROSCI.4908-11.2012>

1051 Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer,
1052 S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as
1053 indicated by pupil dilation. *Hearing Research*, 351, 68–79.
1054 <https://doi.org/10.1016/j.heares.2017.05.012>

1055 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for
1056 advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational*
1057 *Intelligence and Neuroscience.* <https://doi.org/10.1155/2011/156869>

1058 Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related
1059 potential to violations of abstract regularities: A review. *International Journal of*
1060 *Psychophysiology*, 88(2), 109–123. <https://doi.org/10.1016/j.ijpsycho.2013.03.015>

1061 Pashler, H. E. (1998). *The Psychology of Attention*. MIT Press. [http://cognet.mit.edu/book/psychology-of-](http://cognet.mit.edu/book/psychology-of-attention)
1062 [attention](http://cognet.mit.edu/book/psychology-of-attention)

1063 Postle, B. R., Stern, C. E., Rosen, B. R., & Corkin, S. (2000). An fMRI investigation of cortical
1064 contributions to spatial and nonspatial visual working memory. *NeuroImage*, 11, 409–423.
1065 <https://doi.org/10.1006/nimg.2000.0570>

1066 Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007).
1067 Dissociable but inter-related systems of cognitive control and reward during decision making:
1068 Evidence from pupillometry and event-related fMRI. *NeuroImage*, 37(3), 1017–1031.
1069 <https://doi.org/10.1016/j.neuroimage.2007.04.066>

1070 Sauseng, P., Klimesch, W., Freunberger, R., Pecherstorfer, T., Hanslmayr, S., & Doppelmayr, M. (2006).
1071 Relevance of EEG alpha and theta oscillations during task switching. *Experimental Brain*
1072 *Research*, 170(3), 295–301. <https://doi.org/10.1007/s00221-005-0211-y>

1073 Scerra, V. E., & Brill, J. C. (2012). Effect of Task Modality on Dual-Task Performance, Response Time,
1074 and Ratings of Operator Workload. *Proceedings of the Human Factors and Ergonomics Society*
1075 *Annual Meeting*, 56(1), 1456–1460. <https://doi.org/10.1177/1071181312561409>

1076 Schack, B., & Klimesch, W. (2002). Frequency characteristics of evoked and oscillatory
1077 electroencephalic activity in a human memory scanning task. *Neuroscience Letters*, 331(2), 107–
1078 110. [https://doi.org/10.1016/S0304-3940\(02\)00846-7](https://doi.org/10.1016/S0304-3940(02)00846-7)

1079 Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in*
1080 *Cognitive Sciences*, 13(11), 488–495. <https://doi.org/10.1016/j.tics.2009.08.005>

1081 Stensaas, S. S., Eddington, D. K., & Dobelle, W. H. (1974). The topography and variability of the
1082 primary visual cortex in man. *Journal of Neurosurgery*, 40(6), 747–755.
1083 <https://doi.org/10.3171/jns.1974.40.6.0747>

1084 Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual
1085 topography of human intraparietal sulcus. *Journal of Neuroscience*, 27(20), 5326–5337.
1086 <https://doi.org/10.1523/JNEUROSCI.0991-07.2007>

1087 Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α -Band Electroencephalographic
1088 Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target
1089 Detection. *The Journal of Neuroscience*, 26(37), 9494–9502.
1090 <https://doi.org/10.1523/JNEUROSCI.0875-06.2006>

1091 Tobyne, S. M., Osher, D. E., Michalka, S. W., & Somers, D. C. (2017). Sensory-biased attention
1092 networks in human lateral frontal cortex revealed by intrinsic functional connectivity.
1093 *NeuroImage*, 162, 362–372. <https://doi.org/10.1016/j.neuroimage.2017.08.020>

1094 Tombu, M. N., Asplund, C. L., Dux, P. E., Godwin, D., Martin, J. W., & Marois, R. (2011). A unified
1095 attentional bottleneck in the human brain. *Proceedings of the National Academy of Sciences of*
1096 *the United States of America*, 108(33), 13426–13431. <https://doi.org/10.1073/pnas.1103583108>

1097 Tune, S., Wöstmann, M., & Obleser, J. (2018). Probing the limits of alpha power lateralisation as a neural
1098 marker of selective attention in middle-aged and older listeners. *European Journal of*
1099 *Neuroscience*, 48(7), 2537–2550. <https://doi.org/10.1111/ejn.13862>

1100 Vogt, F., Klimesch, W., & Doppelmayr, M. (1998). High-Frequency Components in the Alpha Band and
1101 Memory Performance: *Journal of Clinical Neurophysiology*, 15(2), 167–172.
1102 <https://doi.org/10.1097/00004691-199803000-00011>

1103 Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive*
1104 *Sciences*, 7(12), 553–559. <https://doi.org/10.1016/j.tics.2003.10.012>

1105 Watanabe, K., & Funahashi, S. (2014). Neural mechanisms of dual-task interference and cognitive
1106 capacity limitation in the prefrontal cortex. *Nature Neuroscience*, 17(4), 601–611.
1107 <https://doi.org/10.1038/nn.3667>

1108 Wessinger, C. M., Buonocore, Michael H., Kussmaul, C. L., & Mangun, G. R. (1997). Tonotopy in
1109 human auditory cortex examined with functional magnetic resonance imaging. *Human Brain*
1110 *Mapping*, 5(1), 18–25.

1111 Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics*
1112 *Science*, 3(2), 159–177. <https://doi.org/10.1080/14639220210123806>

1113 Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on
1114 Listening Effort Revealed by Pupil Dilation. *Ear and Hearing*, 36(4), 153–165.
1115 <https://doi.org/10.1097/AUD.0000000000000145>.The

1116 Winn, M. B., & Moore, A. N. (2018). Pupillometry Reveals That Context Benefit in Speech Perception
1117 Can Be Disrupted by Later-Occurring Sounds, Especially in Listeners With Cochlear Implants.
1118 *Trends in Hearing*, 22, 233121651880896. <https://doi.org/10.1177/2331216518808962>

1119 Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory Biasing of Visuospatial
1120 Attention Indexed by Retinotopically Specific α -Bank Electroencephalography Increases over
1121 Occipital Cortex. *The Journal of Neuroscience*, 20(6), RC63–RC63.
1122 <https://doi.org/10.1523/JNEUROSCI.20-06-j0002.2000>

1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141

1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162

1163 Supplemental Figures

1164

1165 **Supplemental Figure 1: Grand average ERPs for each combination of WM task (colors), Intervening**

1166 **task (rows), and position in the stimulus sequence.** The modality of stimulus presentation is shown by

1167 solid versus dashed lines. The top-left legend shows the channels averaged to produce ERPs for each

1168 stimulus modality: on the standard 10-20 layout, these were channels Fz, AFz, Cz, F1, and F2 for auditory

1169 stimuli and O1, O2, PO3, PO4, PO7, PO8 for visual stimuli. Axes below the data in each trial phase are

1170 shown for scale.

1171

Supplemental Figure 2: Grand average oscillatory power during WM encoding in the AT (A), AS (B), VT (C), and VS (D) WM conditions. Data are shown as change relative to the pre-trial baseline, in dB. Alpha power strengthened relative to baseline during auditory WM encoding and became suppressed relative to baseline during visual WM encoding. 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.

Supplemental Figure 3: Grand average power spectra during the pre-trial baseline, plotted separately at channel Pz (A) and channel AFz (B). Note the well-defined peak in the alpha (8-13 Hz) frequency range at the parietal electrode site. Theta (4-7 Hz) power was more task-induced and so was less strong during baseline at the frontal electrode site; nonetheless, the power spectrum in panel B flattens around this frequency range at a higher value than at the parietal electrode site. Error clouds represent S.E.M.

θ -Band (4-7 Hz) Activity

1196

1197 **Supplemental Figure 4: Increased theta (4-7 Hz) power in conditions with an Intervening task. A,**
 1198 Topographies of theta power when the Intervening task was AT, shown to illustrate the fronto-central
 1199 concentration of theta power. **B-E**, Left panels show time-frequency responses at a fronto-central
 1200 electrode site (AFz) relative to the pre-trial baseline. Within each WM task, scalp topographies to the
 1201 right show the results of cluster-based permutation tests comparing the indicated Intervening task
 1202 conditions. T-values from cluster formation are plotted on the green-purple axis, as in some cases, outlier
 1203 participants with Supplemental Figure 3.12 (Continued): anomalously strong theta power changes
 1204 exaggerated the apparent differences in the grand average time-frequency plots. Asterisks signify
 1205 membership in a significant cluster ($p < 0.05$). For each comparison, only the strongest time-channel
 1206 cluster is shown, although some comparisons did yield multiple significant clusters. If the strongest
 1207 cluster identified did not reach significance, the topoplot is nonetheless restricted to the time region of

1208 that cluster. If no clusters were identified, the topoplot time range spans the entire WM retention window
1209 (time shown as N.A.).

A **B** **C**

Working Memory Task

Encoding

Auditory or Visual

Probe

Auditory or Visual

Encoding vs. Probe

Temporal: "Different"
Spatial: "Same"

D

Intervening Task

Auditory-Temporal

Interval 1 or 2 longer?
"2"

Auditory-Spatial

Sound B left or right of A/C?
"Right"

A **B**

A**No Int.****AT Int.****AS Int.**

WM Task

AT VT

AS VS

B**Differential effect of Int. tasks**Int. Task: *AT — None*Int. Task: *AS — None***Difference Plots**

Intervening Task

minus

No Intervening Task

Pupil Diameter Difference (z)

*P2 topographies
(190-220 ms)*

C **Modality Main Effect**

D **Domain – Int. Task Interaction**

Auditory WM

Visual WM