1 **Similarity in sensory modality and information domain impair**

2 **processing in a dual-task context: Evidence from behavior,**

3 **pupillometry, and EEG**

4
5 Justin T. Fleming[1], J. Michelle Njoroge[2], Abigail L. Noyce[3], Tyler K. Perrahione[1,2], and Barbara

6 G. Shinn-Cunningham[3]

7

8   1. Speech and Hearing Bioscience and Technology Program, Harvard University,

9   Cambridge, MA

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

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

12

13 Corresponding author:

14   Justin T. Fleming

15   164 Pillsbury Dr. SE

16   Minneapolis, MN 55455

17   jtf@umn.edu

18

19

20

21

22

23

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

70   maps of space are found throughout the visual processing pathway (Silver & Kastner, 2009;

71   Stensaas et al., 1974; Swisher et al., 2007). At various stages along this pathway, visual neurons

72   are sensitive to particular spatial features, such as edges, object orientation, spatial frequency, and

73   texture (Hubel & Wiesel, 1959; Issa et al., 2000). Auditory spatial information, on the other hand,

74   must be indirectly computed based on interaural cues. The peripheral auditory system is instead

75   tuned to temporal information, with auditory nerve fibers capable of phase-locking to oscillating

76   sound pressure waves (Dynes & Delgutte, 1992). This temporal sensitivity allows the auditory

77   system to be organized around sound frequency (Da Costa et al., 2011; Merzenich & Brugge, 1973;

78   Wessinger et al., 1997), and for perceptual judgement of inter-stimulus intervals to be much more

79   precise in audition than vision (Lhamon & Goldstone, 1974). Reflecting this temporal

80   specialization of the auditory system, the auditory-biased LFC network can be recruited by visual

81   inputs if the task requires temporal information processing. Similarly, the visual-biased LFC

82   network is recruited by auditory inputs if spatial information processing is required (Michalka et

83   al., 2015). This flexible allocation of resources allows information to be processed by a network

84   geared to the complementary strengths of each sensory modality.

85   Given what is known about recruitment in these sensory-biased cortical networks, in the

86   present study we tested the prediction that dual-task interference will depend on whether the two

87   tasks share the same sensory modality (auditory vs. visual) and information domain (temporal vs.

88   spatial), thus causing the tasks to rely upon processing within the same cortical network. We

89   developed a dual-task interference paradigm featuring WM and "Intervening" tasks. The WM task

90   required participants to remember temporal or spatial information about a set of auditory or visual

91   stimuli, while the Intervening task – presented during WM retention – required participants to

92   make an immediate perceptual judgment about either the timing or spatial locations of auditory

93   stimuli. This design was inspired by "multiple resource theory," which posits that two tasks will

94   interfere with each other when they draw on the same pool of neural resources (Navon & Gopher,

95   1979; Nickerson, 1980; Wickens, 2002). This paradigm allowed us to observe patterns of

96   interference as participants engaged a WM network to encode sensory information in memory,

97   performed a real-time perceptual task that required either increasing the load on the active network

98   or switching networks, and finally retrieved the information stored in WM.

99   To gain insights about the temporal dynamics of dual task interference in this paradigm,

100   we combined behavioral measures with two temporally sensitive indexes of cognitive processing:

101   pupillometry and electroencephalography (EEG). Task-evoked pupil dilations dynamically reflect

102   task difficulty or the participant's effort level during task performance (Causse et al., 2016;

103   Gilzenrat et al., 2012; Murphy et al., 2011; Winn et al., 2015). EEG signals reveal different stages

104   of processing through event-related potentials (ERPs) that are phase locked to sensory events

105   (Luck, 2012), and more loosely coupled changes in oscillatory power in stereotypical frequencies

106   that have been linked to different cognitive states (Ward, 2003). These techniques provide multiple

107   viewpoints into dual-task interference with high temporal precision, but importantly, they lack the

108   spatial resolution necessary to directly implicate the complementary sensory-biased cortical

109   networks. Nonetheless, the WM tasks in this study were derived from the tasks Michalka et al.

110   (2015) used to characterize the complementary LFC networks. Thus, we can say with relative

111   confidence that the specific temporal and spatial WM tasks used in the current paradigm

112   differentially engaged these networks.

113   In conditions with greater interference, we expected to observe the following phenomena:

114   increased behavioral errors, both in WM retrieval and auditory perceptual judgments in the

115   Intervening task; larger maximum pupil dilations; reduced ERP amplitudes, reflecting the strain

116 on neural resources needed to process the stimuli; and differing oscillatory signatures. In

117 considering neural oscillations, we were specifically interested in the theta (4-8 Hz) and alpha (8-

118 13 Hz) frequency bands, both of which have been implicated in task switching and WM

119 maintenance (Bastiaansen et al., 2002; Cunillera et al., 2012; López et al., 2019; Obleser et al.,

120 2012; Sauseng et al., 2006).

121   In brief, convergent evidence from behavior, pupillometry, and EEG indicated that

122 interference was generally elevated when the WM and Intervening tasks relied on the same

123 modality- and domain-biased attention and WM network. When the sensory modality differed

124 between tasks, behavioral and autonomic signatures of interference were weak and

125 indistinguishable based on WM domain. In these conditions, however, ERP amplitudes and alpha

126 oscillations uncovered subtler patterns of domain-based interference. These results indicate that

127 switching between complementary attention and WM networks comes at a lower cost than loading

128 processing onto a single network.

129

130

131 **Materials and Methods**

132 *Participants*

133   Twenty-three healthy young adults completed all experimental procedures. Data from three

134 participants was removed due to excessive noise in the pupillometry or EEG data, making for a

135 final sample of twenty individuals (13 female, mean age 20.9 years, standard deviation 2.3 years).

136 One additional participant was excluded from only the time-frequency analyses due to anomalous

137 high-frequency noise in their EEG data. All participants had normal or corrected-to-normal visual

138 acuity and no reported colorblindness. Participants who wore glasses for visual correction and

139   could not use contact lenses were excluded to avoid potential artifacts in the pupillometry data.

140   All participants had clinically normal hearing, defined by tone detection thresholds below 20 dB

141   HL at octave frequencies between 250 Hz and 8 kHz, as confirmed by an audiometric screening.

142   Participants gave written informed consent and were compensated for their participation. All study

143   procedures were approved by the Boston University Charles River Campus Institutional Review

144   Board.

145

146   *Experimental Setup*

147   The experiment was conducted in a darkened, electrically shielded, sound-treated booth.

148   Participants were seated comfortably with their chin resting on a desk-mounted head support (SR

149   Research). A Benq 1080p LED monitor (27-inch diagonal, 120 Hz refresh rate) was positioned in

150   front of the participant at approximately 65 cm distance. The monitor was set to 3% of its

151   maximum brightness level to prevent eye fatigue and pupil diameter saturation. An EyeLink 1000

152   Plus eye tracking system was placed on the desk just below the display for measurement of pupil

153   diameter. Six free-field loudspeakers (KEF E301) were mounted in an arc around the participant

154   at a distance of 1.5 m. Five of the loudspeakers were equally spaced in azimuth at ±90°, ±45°, and

155   0° relative to midline; the sixth was placed immediately to the left of the central loudspeaker, at

156   −4° azimuth, and used only in the Intervening tasks (Fig. 1A). All six loudspeakers were positioned

157   at approximately 5° elevation relative to the horizontal plane of the eyes to reduce obstruction by

158   the visual display. Auditory stimulus presentation was handled using an RME Fireface UCX

159   soundcard.

160
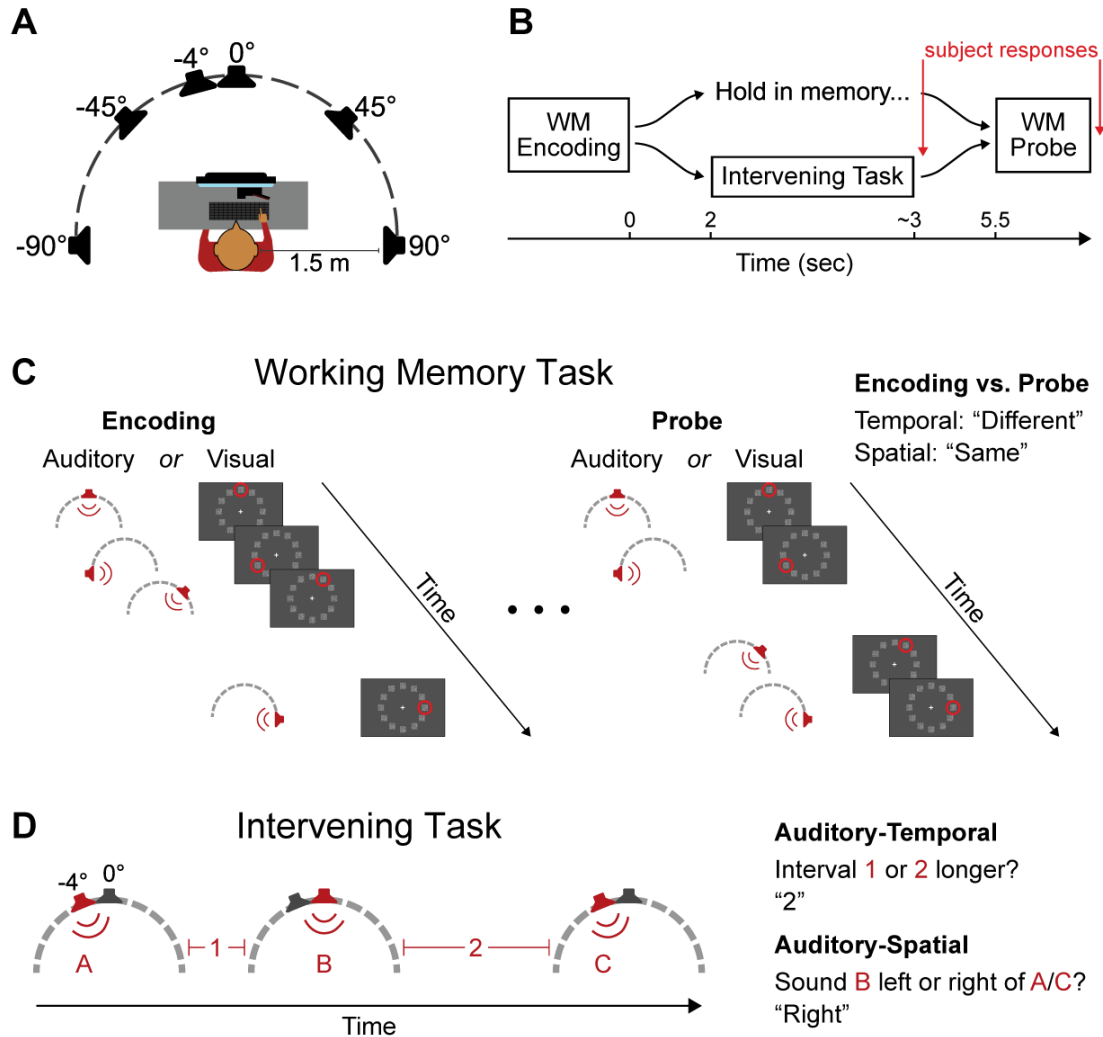
Figure 1: Experimental setup and task design. A, Overhead depiction of the experimental setup. All the loudspeakers except the one at −4° azimuth were used for the WM tasks, while only the loudspeakers at −4° and 0° were used for the Intervening tasks. **B**, Overall dual-task structure. The time base 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

176   and registering behavioral responses. To ensure synchrony of event triggers (e.g. trial starts,

177   stimulus presentation) between the pupillometry and EEG data, triggers were output through the

178   S/PDIF channel on the soundcard, converted to TTL pulses using a custom converter box, and

179   written simultaneously into the EEG and pupillometry data files. Experiment control was carried

180   out using custom MATLAB software, and visual stimulus presentation was implemented using

181   the Psychtoolbox package (Brainard, 1997).

182

183   *Task and Experimental Design*

184   Participants performed a dual-task paradigm, comprising a working memory (WM) task

185   and an Intervening task (Fig. 1B). Each trial started with a 1.5-second baseline period, followed

186   by the presentation of a sequence of four auditory or visual stimuli to be encoded in WM. Each

187   stimulus was presented at one of five (auditory) or twelve (visual) locations, and each inter-

188   stimulus interval in the sequence was randomly set to be either short or long (more details below).

189   The WM task could be either temporal or spatial, yielding four total WM task conditions: auditory-

190   temporal (AT), auditory-spatial (AS), visual-temporal (VT), and visual-spatial (VS). When the

191   WM task was temporal, participants were instructed to remember the pattern of inter-stimulus

192   intervals (i.e., the rhythm), regardless of spatial locations. When the WM task domain was spatial,

193   participants had to remember the locations of the stimuli, regardless of order or timing (Fig. 1C).

194   In all but the AS WM task (see below), both the locations and intervals could change between the

195   encoding and probe sequences, but participants were instructed to ignore changes in the unattended

196   domain.

197   Participants retained stimulus information in WM for 5.5 seconds, after which a four-

198   stimulus probe sequence was presented in the same sensory modality as the encoded sequence.

199   Participants compared the encoded and probe sequences and made a same-different judgment on

200   the remembered domain (temporal or spatial). After the conclusion of the probe stimulus,

201   participants had 1.5 sec to indicate whether the encoded and probe sequences were the same (by

202   pressing "1" on the keyboard) or different (by pressing "0"). Each block contained an equal number

203   of same and different trials, ordered randomly. Participants maintained fixation on a small black

204   cross (0.41° visual angle) at the center of the display throughout the trial.

205   Pilot testing with earlier versions of these WM tasks revealed necessary adjustments to

206   prevent ceiling or floor effects in some conditions. For participants to perceive the different inter-

207   stimulus intervals equally well, a larger separation was needed for the visual stimuli (200 and 580

208   ms) than the auditory stimuli (200 and 340 ms). Conversely, the visual-spatial task was too easy

209   with only five stimulus locations, so the number of potential visual locations was increased to 12.

210   Finally, participants struggled to perform the AS WM task when stimulus timing was variable;

211   therefore, both encoding and probe stimuli in this condition were presented isochronously at the

212   longer inter-stimulus interval.

213   On some trials, participants also performed an Intervening task during the WM retention

214   period. This task was always auditory to allow pupil diameter to be measured in the absence of

215   any visual stimulation, but like the WM tasks, it could be either temporal or spatial (AT or AS; see

216   Fig. 1D). The stimulus structure was the same for the temporal and spatial variants. Starting 2 sec

217   after the offset of the final stimulus in the WM encoding phase, a sequence of three auditory stimuli

218   was presented. These stimuli were white noise bursts, acoustically distinct from the stimuli used

219   in the auditory WM tasks (tone complexes). One of the two intervals between the stimuli was

220   randomly chosen to be slightly longer than the other. The precise intervals were jittered on each

221   trial, with an average interval duration of 460 ms and a 90 ms average difference between the two

222   intervals. The sounds were presented from the two near-frontal loudspeakers (−4° and 0° azimuth).

223   The first stimulus played from one of these loudspeakers, chosen randomly and with equal

224   probability; the second stimulus was played from the other loudspeaker, and the third was played

225   from the same location as the first. In the temporal Intervening task, participants judged whether

226   the first or second inter-stimulus interval was longer. In the spatial Intervening task, participants

227   were asked to determine whether the second sound was to the left or right relative to the first and

228   third sounds.

229   This stimulus design allowed physically identically auditory stimuli to be used for the

230   spatial and temporal Intervening task conditions. However, it did introduce an asymmetry between

231   conditions in the amount of information required to do the task. In the temporal Intervening task,

232   participants needed to attend all three stimuli in order to compare the two inter-stimulus intervals,

233   whereas in the spatial Intervening task, participants were often able to make their judgment by the

234   second auditory stimulus by comparing its location to the first. This could result in a longer period

235   of increasing pupil size in the temporal task, leading to larger peak pupil diameter. However, our

236   analyses are primarily focused on comparing the four WM conditions *within* each Intervening task,

237   as behavioral accuracy was unequal between the two tasks. By deemphasizing across-task

238   comparisons, we sought to minimize the impact that potential differences in behavioral strategy

239   could have on our main conclusions.

240   Participants registered Intervening task responses with a keypress immediately after the

241   last Intervening task stimulus. Thus, neural signatures of motor planning and execution may be

242   present in the EEG data near the end of the WM retention phase. However, we deemed this less

243   detrimental than asking participants to hold their Intervening task responses until the end of the

244   trial, as such a design would increase the WM load in trials with an Intervening task and obfuscate

245   comparison with the no Intervening task conditions. Furthermore, any motor components in the

246   EEG data should be present whenever there was an Intervening task, allowing for fair comparison

247   between the different combinations of WM and Intervening tasks.

248   Trials were grouped into blocks of 20, with the WM and (if present) Intervening task

249   conditions held constant throughout the block. At the start of each block, an instruction screen

250   indicated the sensory modality and relevant domain (temporal or spatial) of the upcoming trials.

251   Participants were allowed to take untimed breaks between blocks. Participants performed one

252   block of each WM and Intervening task combination before any conditions were repeated, and the

253   same condition was not allowed to repeat in adjacent blocks. In total, participants performed 40

254   trials of each combination of WM task modality, WM task domain, and Intervening task condition.

255   Each complete dataset required three separate visits to the lab. The first session was

256   reserved for consent, subject screening, and task practice. Participants practiced each variant of

257   the WM and Intervening tasks in isolation until they understood the procedure, then three to five

258   example trials of the full dual-task paradigm. Data collection for the actual experiment occurred

259   in the two subsequent sessions, which were split based on the sensory modality of the WM tasks;

260   auditory and visual WM tasks were performed on separate days. Session order was randomized

261   and counterbalanced across participants.

262

263   *Stimulus Details*

264   For the visual WM tasks, 12 stimuli were arranged in a circle centered on the fixation cross

265   and shown on a constant dark grey background (2.51 cd/m^2). Each stimulus was a square patch of

266   visual noise, subtending $2.86°$ of visual angle and composed of a 30 x 30 grid of smaller squares.

267   Each of these smaller squares was filled with a greyscale color between black and white, such that

268   the average luminance across the patch was 5.85 cd/m^2. The angular spacing between each patch

269   was 30°, and the entire stimulus circle subtended 21.59° of visual angle. To equate display

270   luminance and structure across tasks, these visual stimuli remained present but static throughout

271   the auditory WM and Intervening tasks. To employ these stimuli in the visual WM tasks, the

272   luminance of each small square in a given patch could be resampled; this made the visual patch

273   appear to jitter without changing the average luminance across the patch.

274   For the auditory WM tasks, each stimulus was a 50-ms tonal chord consisting of 3

275   harmonically unrelated complex tones (fundamental frequencies of 422, 563, and 670 Hz); each

276   complex was made up of its first nine harmonics with equal amplitude. The same tone complex

277   was used for all auditory stimuli in the WM tasks. WM task stimuli were only presented from the

278   loudspeakers at ±90°, ±45°, and 0° azimuth. For the Intervening tasks, which were always auditory,

279   the stimuli were one of five pre-generated, 50-ms bursts of noise bandpass filtered between 100

280   and 10,000 Hz. Identical noise tokens were used for all three stimuli within each Intervening task

281   sequence. These WM and Intervening task stimuli were chosen to be distinct, such that stimuli

282   were clearly different between the two tasks. Both types of stimuli were relatively broadband, thus

283   ensuring they provided rich and robust spatial localization cues. All auditory stimuli were ramped

284   on and off with a 5 ms cosine-squared ramp to avoid spectral splatter and transient onset and offset

285   artifacts.

286   In the visual WM tasks, the first stimulus to change in the encoding and probe sequences

287   was always the one at top-center (12 o'clock). Similarly, in all auditory WM sequences, the first

288   sound was presented from the central loudspeaker. This was done to equate the number of items

289   stored in WM across the spatial and temporal WM tasks. With four stimuli in each sequence, there

290   were three intervals to remember for the temporal tasks, and so the first stimulus location was held

291   constant such that only three locations needed to be remembered for the spatial tasks.

292

293   *Behavioral Data Analysis*

294   The primary behavioral metrics in this study were the error rates on the WM and

295   Intervening tasks. Reaction time estimates could not be made because 1) the position in the probe

296   sequence at which a decision could be made differed based on whether the correct response was a

297   "same" (final stimulus) or "different" (often earlier); and 2) within "different" trials, any of the

298   second, third, or fourth stimuli could contain the change.

299   Statistical outcomes of behavioral performance were analyzed using logistic mixed effects

300   regression models. For the WM task, the model included fixed effect terms for WM task modality,

301   WM task domain, and Intervening task, and random effects terms to capture participant-specific

302   intercepts for all three predictor variables. In Wilkinson notation, the model was specified as:

$$logit(Error\ Rate) \sim Modality_{WM} * Domain_{WM} * Int$$

$$+ (1 + Modality_{WM} + Domain_{WM} + Int \mid ID)$$

305   Due to the fundamental differences between the four WM conditions, differences in retrieval

306   between these conditions were not of primary interest. Instead, we focused post-hoc analyses on

307   how retrieval in each WM condition was impacted by modality and domain similarity with the

308   Intervening task. To this end, post-hoc comparisons were only conducted within each WM task

309   condition (e.g., examining differences in AT WM retrieval between the AT, AS, and No

310   Intervening task conditions). Similarly, for Intervening task performance, we held the Intervening

311   task fixed and compared error rates between the four WM task conditions. This analysis required

312   us to compare across levels of both WM modality and WM domain; to facilitate this, the

313 Intervening task model was constructed with these factors collapsed into a single predictor

314 variable, WM Condition:

315 $$logit(Error\ Rate) \sim WM\ Condition * Int + (1 + WM\ Condition + Int \mid ID)$$

316 To investigate main effects and interactions at the group level, the coefficients from these

317 models were supplied to a 3-way (WM task) or a 2-way (Intervening task) repeated measures

318 ANOVA. Contrasts were initially treatment coded with base levels set as "auditory" for WM

319 modality, "temporal" for WM domain, and "none" for Intervening task – these choices had no

320 impact on the outcome of the group-level ANOVA. However, pairwise post-hoc testing was

321 conducted by cycling which level was considered "baseline" for each factor until a β weight (and

322 corresponding p-value) could be extracted for each necessary pair of levels. All post-hoc tests were

323 Bonferroni-Holm corrected for multiple comparisons.

324

325 *Pupillometry data collection and analysis*

326 Both pupils were tracked with a combination of the corneal reflection and a fit of the pupil

327 border using SR Research algorithms. Prior to the start of each experimental session, eye position

328 measurements were calibrated based on five fixation points (display center, ±20° azimuth on the

329 horizontal plane, and ±10° elevation on the median plane). This calibration was validated at the

330 center position prior to each trial, and at all five points at the start of each block. Whenever eye

331 position was offset from the calibration by an average of greater than 4°, the full five-point

332 calibration was repeated. The experimenter monitored gaze position throughout the experiment; if

333 at any point during the trial the participant's gaze deviated substantially from 0° in either azimuth

334 or elevation, the experimenter immediately stepped in to ensure that the participant maintained

335 center fixation moving forward.

336   A custom MATLAB analysis pipeline and wrapper GUI were used to prepare pupil data

337   for statistical analysis. First, trials were split into baseline and trial windows, where the trial

338   window spanned the start of the WM encoding sequence through the end of the WM retention

339   phase. Next, blinks were automatically detected based on instantaneous position, velocity, and

340   acceleration thresholds. An experimenter manually reviewed the data, and using the GUI, adjusted

341   blink thresholds or manually marked additional blink segments as needed. Blinks and other marked

342   segments of noisy data were replaced with a linear interpolation between the average of the three

343   samples preceding and following the blink. When blinks occurred at the beginning of the trial

344   window, a linear fit was made to the five samples following the blink, and this fit was back-

345   projected through the blink segment. The opposite procedure was used for blinks falling at the end

346   of the trial window. Trials in which more than 25% of the data was made up of rejected segments

347   were automatically excluded from further analysis. When data from both eyes were available, the

348   two traces were averaged. To produce the final output, the traces were concatenated, Z-scored,

349   then split back into trials and trial windows. This procedure eschewed absolute pupil size measures

350   in favor of values that were individually normalized for each participant, so only relative pupil

351   diameter between conditions is interpreted.

352   Statistical testing for differences between conditions in the pupil time courses was carried

353   out using non-parametric permutation tests. First, at each time point, the difference between two

354   conditions was assessed parametrically using a paired T-test across the individual participant

355   average data. Next, 2000 iterations of randomly shuffling the condition labels for each participant's

356   data were performed; these randomly labeled samples were used to recompute paired T-tests at

357   each time point, generating a null distribution of 2000 such values. Finally, at each time point, the

358   significance level of the actual difference was determined by calculating the percentage of the null

359   distribution with a T-value equal to or larger than the T-value from the actual data. Significant

360   differences were only considered reliable if the p-value fell below 0.05 for a minimum of 15

361   consecutive samples.

362

363   *EEG Data Analysis*

364   EEG analyses were carried out using the FieldTrip package in MATLAB (Oostenveld et

365   al., 2011). For event-related potential (ERP) analyses, EEG preprocessing comprised the following

366   steps: read in the continuous data one channel at a time and immediately downsample to 256 Hz;

367   re-reference the data to the average of two electrodes placed on the mastoids, bandpass filter

368   between 0.5 and 20 Hz (zero-phase FIR filter, transition width of 0.2 Hz, order of 9274); manually

369   identify and remove segments containing muscle artifacts; perform an independent components

370   analysis (ICA) to project out blinks and saccadic eye movements; epoch the data from 100 ms

371   before to 500 ms after each individual auditory or visual stimulus (timing differences precluded

372   whole-trial averaging); reject any epochs in which the data exceeded a 100 µV peak-to-peak

373   threshold; and baseline correct by subtracting off the mean of the first 100 ms of each epoch.

374   For time-frequency analyses, a similar preprocessing pipeline was used, but with a few key

375   differences. First, the low-pass filter cutoff was raised to 80 Hz. Second, participant average ERPs

376   (recomputed with the new filter cutoffs) were subtracted from the data in the time domain

377   immediately prior to epoching; this step served to minimize phase-locked evoked contributions to

378   the time-frequency response. Subtracted ERPs were specific to each trial phase (i.e. encoding,

379   retention, and probe), WM and Intervening task condition, and position in the stimulus sequence.

380   Third, the data was split into whole-trial epochs, spanning the baseline period through the final

381   stimulus in the probe sequence, instead of shorter individual stimulus epochs. The continuous

382 Morlet wavelet transform (wavelet width of 5 cycles in 1 Hz steps) was used to obtain the power

383 spectra of each trial. Prior to wavelet analysis, the signal was padded to avoid edge artifacts. This

384 was done by copying the first and last 5 seconds of the epoch, reflecting each copy on the time

385 axis, then appending them to the beginning and end of the signal, respectively. Finally, the data

386 was split into the three key trial phases: WM encoding, WM retention/Intervening task, and WM

387 probe.

388   We extracted time courses from the resulting power spectra at each channel in the theta (4-

389 8 Hz) and alpha (8-13 Hz) frequency bands. To produce theta time courses, power estimates

390 centered on 4, 5, 6, and 7 Hz (1 Hz bandwith) were averaged for each participant. Individual

391 differences in the peak frequency of alpha oscillations are well-established (Klimesch et al., 1999).

392 Thus, for alpha time courses, the individual alpha frequency for each participant was defined as

393 the frequency at which the absolute value of the alpha change relative to baseline was maximal

394 during WM retention (the direction of alpha change was found to flip based on WM modality in

395 the present study). Alpha power time courses were reported as the power at this frequency averaged

396 with the power 1 Hz above and below the individually defined alpha frequency.

397   Similar to the pupillometry data, EEG data were analyzed using non-parametric

398 permutation testing, this time implemented using FieldTrip software. To control the multiple

399 comparisons problem with high-dimensional EEG data, we used cluster-based permutation testing

400 to analyze ERPs and oscillatory power time courses. Paired t-tests were first conducted at each

401 time-channel data point. Those comparisons reaching a significance threshold (generally $p < 0.01$)

402 were grouped into clusters of points contiguous in time and within 40 mm scalp distance of one

403 another. The "mass" of this cluster was then computed as the sum of T-scores across all member

404   points. Finally, this cluster mass was compared against a null distribution of clusters formed by

405   2000 random permutations of the condition labels.

406

407   **Results**

408   *Intervening Task Performance*

409   Performance accuracy on the Intervening tasks depended on both the Intervening task

410   condition and the modality and domain of the information participants were holding in WM (Fig.

411   2). These effects were analyzed using a logistic mixed-effects model with fixed-effect terms of

412   Intervening task (AT or AS) and WM task (AT, AS, VT, or VS). Random-effects terms included

413   by-participant intercepts for both fixed effects. Group-level significance of the fixed effects and

414   their interaction were tested in a Type-III ANOVA, with p-values based on the Satterthwaite

415   approximation for degrees of freedom. This ANOVA revealed significant main effects of

416   Intervening task ($\chi^2 = 32.7$, df = 1, p = $1.07 \cdot 10^{-8}$) and WM task ($\chi^2 = 17.6$, df = 3, p = $5.29 \cdot 10^{-4}$),

417   as well as a significant interaction between these factors ($\chi^2 = 24.1$, df = 3, p = $2.44 \cdot 10^{-5}$). The

418   main effect of Intervening task indicates that performance was better overall on the AS Intervening

419   task than the AT Intervening task. This difference simply reflects a limit on the difficulty of the

420   AS Intervening task imposed by how close together the two loudspeakers could be placed. Since

421   performance differed widely between the two Intervening tasks, we focused pairwise post-hoc

422   tests on conditions in which the Intervening task was the same; differences between these

423   conditions could only be caused by the type of information participants were holding in WM.

424
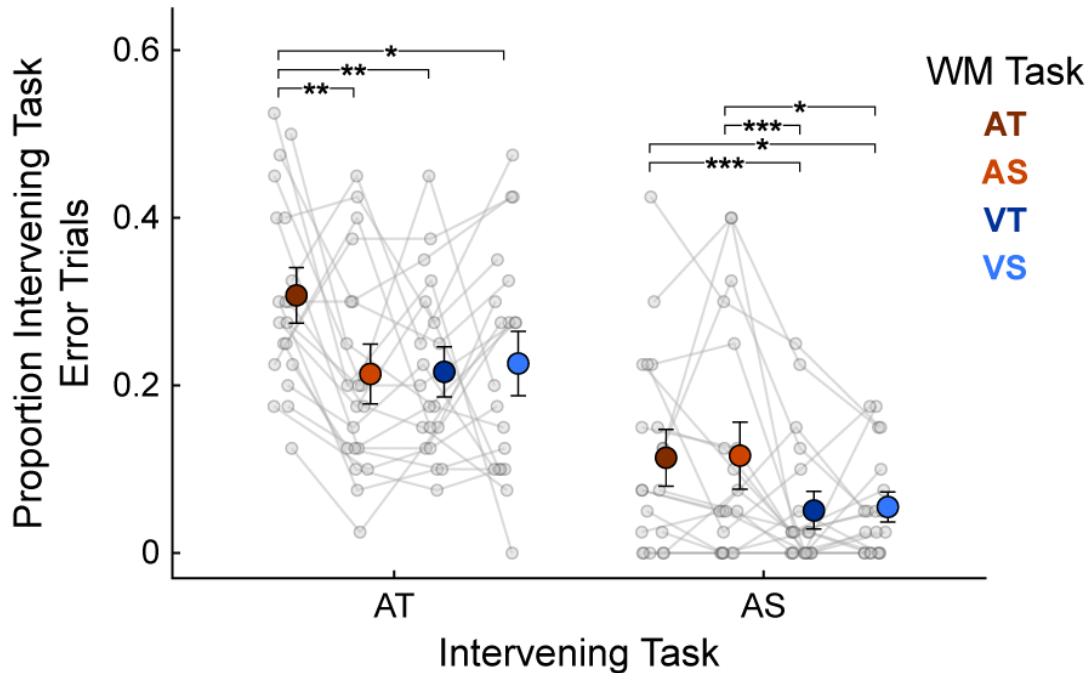
Figure 2: Intervening task error rates for each combination of WM task and Intervening task. Grey points represent individual participants, colored circles represent means, and error bars represent S.E.M. Chance performance is at an error trial proportion of 0.5. Asterisks indicate significant contrasts 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.

430

431   As a means of post-hoc testing, we report the significance of contrasts in the mixed-effects

432   model coding for each pair of WM tasks, with p-values adjusted for multiple comparisons using

433   the Bonferroni-Holm correction. Participants made significantly more errors on the AT

434   Intervening task when the information being held in WM was also AT, as compared to all the other

435   WM conditions (AT vs. AS, p = 0.006; AT vs. VT, p = 0.003; AT vs. VS, p = 0.03; Fig. 2, left).

436   Accuracy on the AS Intervening task appeared to depend primarily on the sensory modality of the

437   WM task, with participants making significantly more errors when both tasks were auditory than

438   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 =

439    $5.50 \cdot 10^{-5}$; AS vs. VS, p = 0.021; Fig. 2, right). Note that error rate floor effects were apparent on

440   the AS Intervening task, which may have limited our ability to detect additional interference effects

441 based on WM domain. Nonetheless, participants generally committed more errors on the

442 Intervening task when it matched the modality – and in the case of the AT Intervening task, the

443 domain – of the information stored in WM.

444

445 *Working Memory Task Performance*

446   Participants were able to retrieve the information stored in WM at better-than-chance levels

447 across all combinations of WM modality, WM domain, and Intervening task (Fig. 3). These factors

448 (including conditions with no Intervening task) were included as fixed effects in a logistic mixed

449 effects model of WM task retrieval errors. Random-effects terms included by-participant intercepts

450 for all fixed effects, and group-level significance of the fixed effects was again tested using a Type-

451 III ANOVA. This ANOVA revealed a significant interaction between WM task modality and

452 domain ($\chi^2 = 206.4$, df = 1, p < 10^{-15}), reflecting overall lower error rates on the AT and VS WM

453 tasks, in which the sensory modality was optimally suited for the information domain of the task

454 (Fig. 3A). The ANOVA also revealed a three-way interaction between WM modality, WM

455 domain, and Intervening task ($\chi^2 = 19.0$, df = 2, p = $7.66 \cdot 10^{-5}$), indicating differential effects of the

456 Intervening tasks depending on WM condition. These effects were explored via post-hoc testing,

457 again restricted to conditions in which the task being measured – in this case, the WM task – was

458 the same, with differences in WM retrieval now attributable only to the Intervening task type.
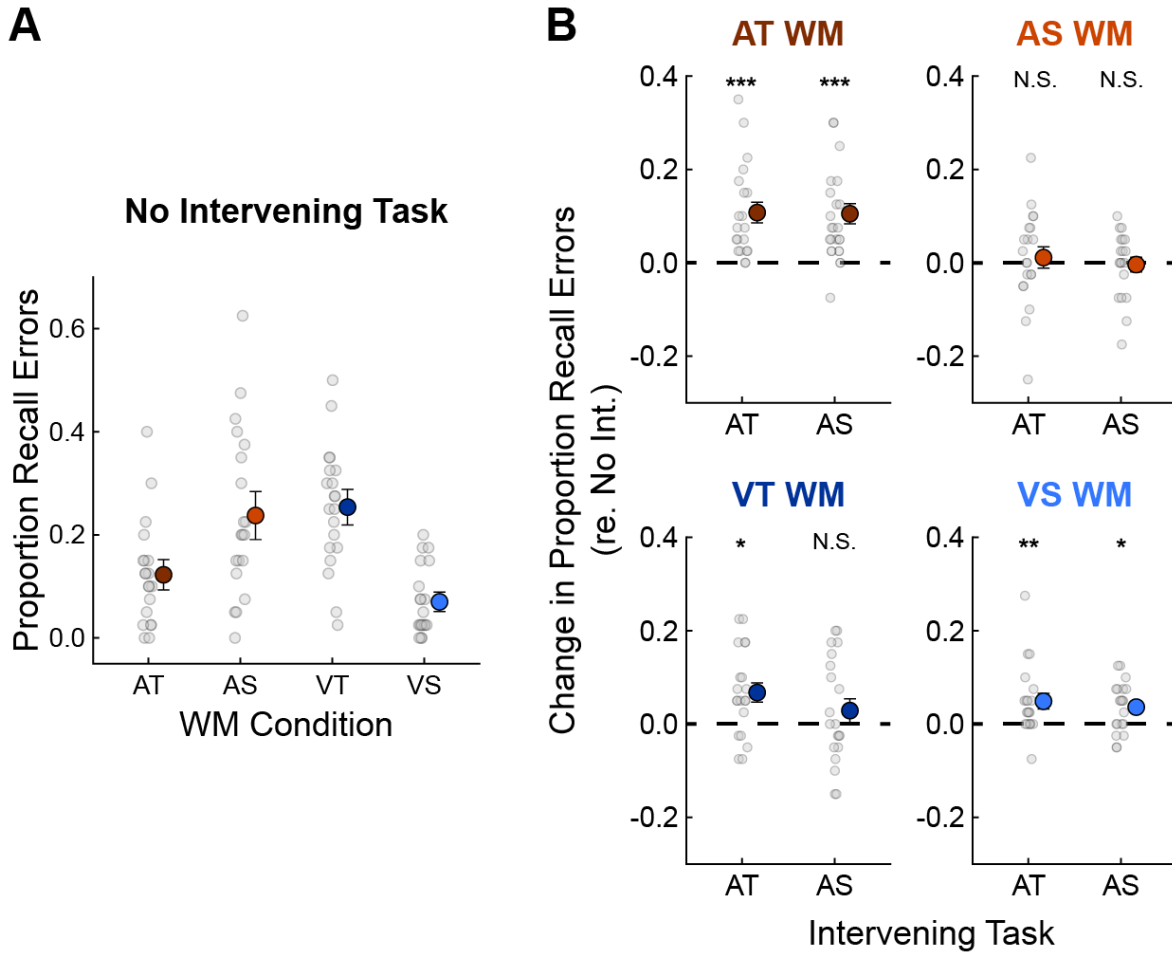
459
460 **Figure 0: Working memory task performance. A**, The proportion of trials on which the WM information
461 was retrieved incorrectly is plotted for conditions without an Intervening task. Chance performance is at
462 0.5. B, The change in retrieval error rate is plotted for each WM and Intervening task combination relative
463 to the no Intervening task conditions. The dashed line at 0 indicates no change in WM retrieval. Post-hoc
464 comparisons were limited to differences between Intervening task conditions within each WM task.
465 Asterisks indicate significant differences from the corresponding no Intervening task conditions (no
466 significant differences were found between the two Intervening tasks). * = $p < 0.05$, ** = $p < 0.01$, *** = p
467 < 0.001 after Bonferroni-Holm correction.
468

469   To visualize effects of the Intervening tasks on WM task accuracy, Fig. 3B shows the

470 difference between error rates when an AT or AS Intervening task was present and the

471 corresponding condition with no Intervening task. When the WM task was AT, both of the auditory

472 Intervening tasks significantly impaired WM retrieval ($p < 10^{-6}$ for both; top-left panel of Fig. 3B).

473 When the WM task was AS, on the other hand, the Intervening tasks had no detectable impact on

474 WM retrieval (top-right panel). We suspect that retaining AT information in WM relied solely on

475 the auditory-temporal WM network, leading to memory interference from the auditory Intervening

476 tasks. In the AS WM task, on the other hand, sound locations to be remembered were likely

477 mapped into a representation in the visual-spatial WM network, protecting this information from

478 auditory interference.

479 The fact that accuracy on the AS WM task was relatively poor might have impaired our

480 ability to detect effects of the Intervening tasks in this condition, as retrieval was closer to chance

481 before the Intervening task was added (Fig. 3A). To explore this possibility, we performed a

482 median split of the AS WM task data in the no Intervening task condition, then tested for

483 differences in the effects of the Intervening tasks between the better and worse performing

484 participants. One might expect that the Intervening tasks would impair retrieval only among the

485 better performers; this group started further from chance performance, resulting in more room for

486 their performance to be affected by the Intervening tasks. However, this was not the case; neither

487 Intervening task significantly altered AS WM task performance for either the better or worse

488 performers, and the effects of the Intervening tasks did not differ between the two groups

489 (uncorrected paired T-tests, all p-values > 0.27).

490 In some cases, visual WM retrieval was also modestly impaired by the auditory Intervening

491 tasks. A small but significant increase in retrieval errors on the VT WM task was detected when

492 the Intervening task was also temporal (p = 0.042; bottom-left panel of Fig. 3.6B). This might

493 reflect domain-based WM interference with both tasks drawing on an auditory-temporal network,

494 although no significant difference was found between the AT and AS Intervening task conditions.

495 In the VS WM condition, both Intervening tasks increased retrieval errors (p = 0.003 for AT

496 Intervening task, p = 0.044 for AS; bottom-right panel). This result was surprising given the lack

497 of modality or domain similarity between the WM and Intervening tasks. However, in this case

498 the errors may reflect a task-switching cost as participants shifted from performing an auditory

499 perceptual judgment to retrieving visual-spatial information from WM.

500

501 *Pupil Dilations*

502   Encoding information in WM and performing the auditory Intervening tasks both led to

503 reliable pupil dilations, shown in Fig. 4. In the absence of an Intervening task, pupil diameter

504 increased throughout presentation of the stimuli to be encoded in WM (negative values on the time

505 axis), peaked about 700 ms after the final WM encoding stimulus (time zero), and then gradually

506 declined throughout WM retention (Fig. 4A, left panel). The onset of the AT and AS Intervening

507 tasks (at 2 sec on the time axis) produced large secondary dilations, the amplitude of which was

508 modulated by the information participants were holding in WM (Fig. 4A, middle and right panels).

509   Statistical differences between the pupil responses were assessed using non-parametric

510 permutation testing, restricted to a time window spanning the WM retention phase (time zero and

511 later). Tests was not conducted during WM task stimulus presentation because 1) the traces were

512 baseline-corrected to zero at the start of this window, and 2) the interstimulus-intervals differed

513 between auditory and visual WM conditions (see Methods), leading to different trace durations. In

514 the absence of an Intervening task, AS WM encoding elicited a markedly larger pupil response

515 than any of the other WM conditions (Fig. 3A). This may have reflected the generally higher

516 difficulty of the AS WM task as compared to the AT WM task. However, while a similar

517 performance difference was found between the VT and VS conditions, no difference in pupil

518 dilations was detected between these conditions. Thus, these results may suggest an asymmetry

519   between the cognitive costs of encoding auditory information into a visual-spatial framework (AS

520   WM) and encoding visual information into an auditory-temporal representation (VT WM).

521
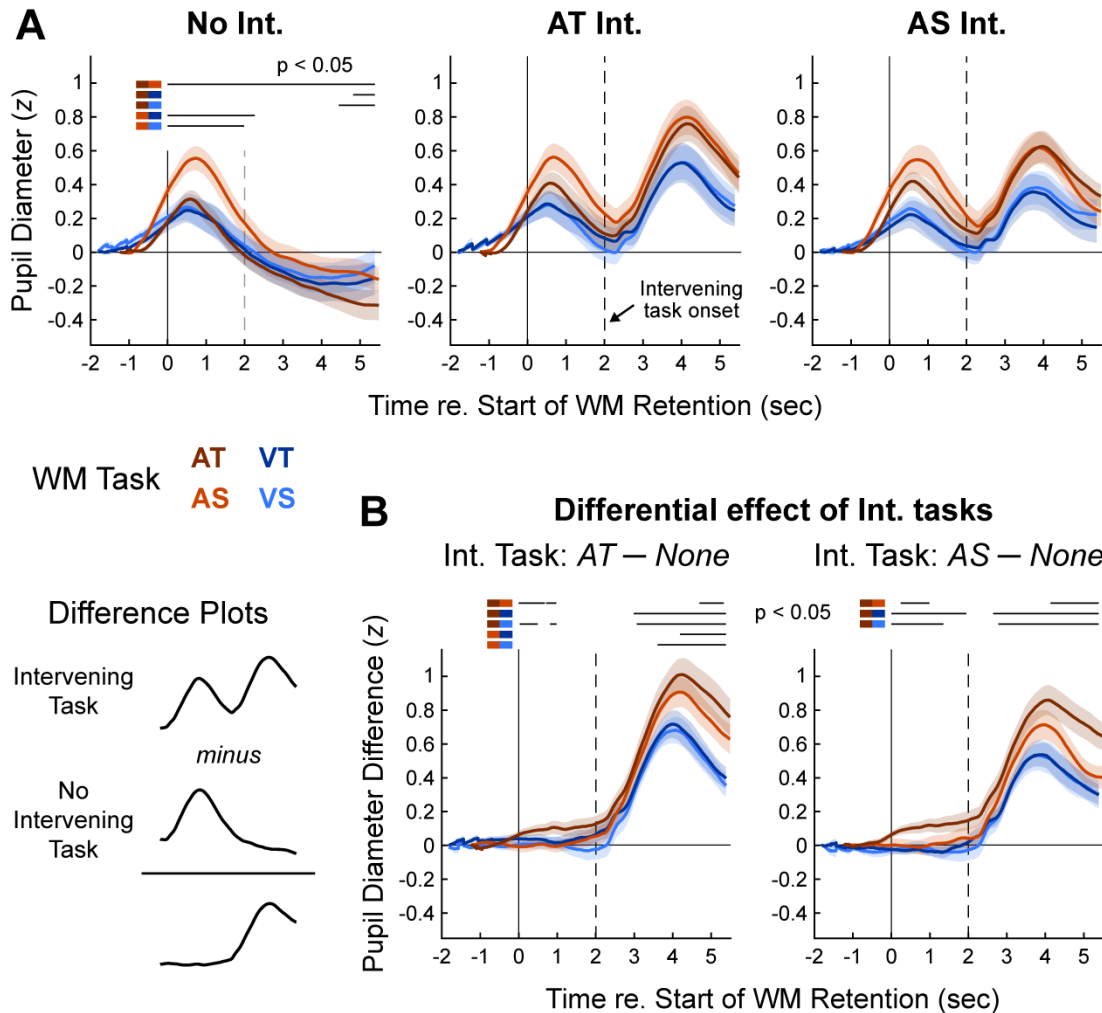
522
523   **Figure 4: Pupillometry results. A**, Grand average z-scored pupil responses in the No Intervening task
524   condition (left) and elicited by physically identical stimuli in the AT (middle) and AS (right) Intervening
525   tasks. Solid vertical lines represent the end of WM encoding and the start of WM retention, while dashed
526   vertical lines indicate the onset of the Intervening task. Error clouds represent S.E.M. Horizontal black lines
527   above the traces represent time regions of significant difference in permutation testing, with the two WM
528   conditions being compared indicated by colors next to each significance line. **B**, The difference between
529   pupil responses elicited in conditions with an Intervening task and responses in the corresponding WM
530   conditions with no Intervening task.

531

532   Pupil dilations elicited by both auditory Intervening tasks were larger when the information

533   stored in WM was auditory than when it was visual, mirroring the modality-driven interference

534   effects observed in the behavioral data (Fig. 4A, middle and right panels). To more clearly separate

535   the differential impact of the Intervening tasks on pupil diameter, we subtracted corresponding

536   pupil traces with no Intervening task from each pupil trace with an Intervening task (Fig. 4B). As

537   a preliminary analysis, we averaged these residual pupil traces across WM conditions and tested

538   for a difference between the two Intervening tasks. This confirmed that traces elicited by the AT

539   Intervening task were larger overall than those elicited by the AS Intervening task ($p < 0.05$ for all

540   time points between 3.7 sec after WM retention onset and the end of the WM retention window;

541   not shown), matching the higher behavioral error rates on the AT task.

542   Next, we compared the residual pupil traces between the four WM conditions within each

543   Intervening task, similar to our treatment of the Intervening task behavioral data. Differential pupil

544   dilations elicited by both Intervening tasks were larger when the information held in WM was

545   auditory than when it was visual, as can also be observed in the raw traces in Fig. 4A. However,

546   expressing the pupil dilations relative to conditions without an Intervening task revealed the largest

547   differential response when the WM task was AT – significantly larger than when the WM task was

548   AS – for both Intervening tasks. Further, the residual pupil traces were elevated prior to Intervening

549   task onset in the AT WM condition, indicating larger pupil dilations during WM encoding when

550   an Intervening task was upcoming. This suggests a preparatory modulation of task effort, with the

551   AT WM information needing to be encoded with particularly high fidelity in order to survive

552   interference from an auditory Intervening task. These results are broadly consistent with the

553   behavioral data, in which the auditory Intervening tasks tended to be most costly when the WM

554   task was AT. We expect that the AT WM condition maximally loaded auditory-biased attention

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

556   Intervening tasks.

557   In the visual WM conditions, pupil dilations were insensitive to whether the WM task was

558   temporal or spatial. This result mirrors the behavioral data and indicates that interference was

559   generally low when the two tasks were presented in different sensory modalities, even if they were

560   matched in information domain. Importantly, the patterns of pupil dilations we observed cannot

561   readily be explained by differences in WM task difficulty. Average pupil dilations were larger with

562   auditory than visual information held in WM, particularly when the WM task was AT. A pure

563   difficulty-based account would therefore predict that WM retrieval should be worst in the AT

564   condition, when in fact participants performed worst on the AS and VT WM tasks, in which the

565   sensory modality was sub-optimal for the information domain of the task. Instead, the pupil

566   dilations observed here likely reflect patterns of dual-task interference specific to the similarity in

567   modality and domain between the two tasks.

568

569   *Event-related potential (ERP) amplitudes*

570   While the behavioral and pupillometry data revealed consistent effects of the modality and

571   domain of the WM task, the domain of the Intervening task had relatively little impact on the

572   pattern of dual-task interference. One explanation for this could be that the AS Intervening task

573   did not recruit the visual-spatial attention network to the same extent as the AS WM task. The AS

574   Intervening task required only an immediate, relative judgement about sound locations, whereas

575   the AS WM task required participants to remember absolute sound positions by placing them in

576   allocentric reference frame. Thus, both Intervening tasks may have predominantly loaded on the

577   auditory-temporal attention network, resulting in similar behavioral and autonomic signatures of

578   interference. Nonetheless, the AT and AS Intervening tasks did require different types of auditory

579   processing, which could affect patterns of dual-task interference at the neural level; we next

580   investigated this possibility in the EEG data.

581   We first examined event-related potentials (ERPs) elicited by stimulus events in the WM

582   and Intervening tasks. Grand average ERPs elicited by each stimulus event can be seen in

583   Supplemental Fig. 1. ERPs elicited by stimuli in the encoding and probe sequences of the WM

584   task were not systematically modulated by the Intervening tasks. However, ERPs elicited by the

585   Intervening task stimuli – particularly the first sound in the Intervening task sequence – were

586   modulated by the information participants were holding in WM. Unlike ERPs elicited by the

587   auditory WM tasks, the Intervening task ERPs had a somewhat atypical morphology, including

588   weak N1 and strong P2 components. This may have been due to the acoustic differences between

589   the WM and Intervening task stimuli, or modulation of ERP morphology due to the fact that

590   participants were holding information in WM. Regardless, both the modality and domain of the

591   information in WM affected the amplitude of the P2 component of the Intervening task onset ERPs

592   (Fig. 5, panels A and B). To quantify these effects, P2 amplitudes were calculated for each

593   participant as the average of the ERP waveform between 170 and 230 ms post-stimulus across a

594   cluster of fronto-central electrode sites (Fz, FCz, Cz, FC1, and FC2 on the standard 10-20 layout).

595   These P2 amplitudes were then modeled using a 3-way ANOVA with explanatory factors of WM

596   modality, WM domain, and Intervening task. This analysis revealed a significant main effect of

597   WM modality ($F = 7.27$, $df = 1$, $p = 0.014$) and a significant interaction between WM domain and

598   Intervening task ($F = 10.83$, $df = 1$, $p = 0.004$), both of which will be examined in more detail
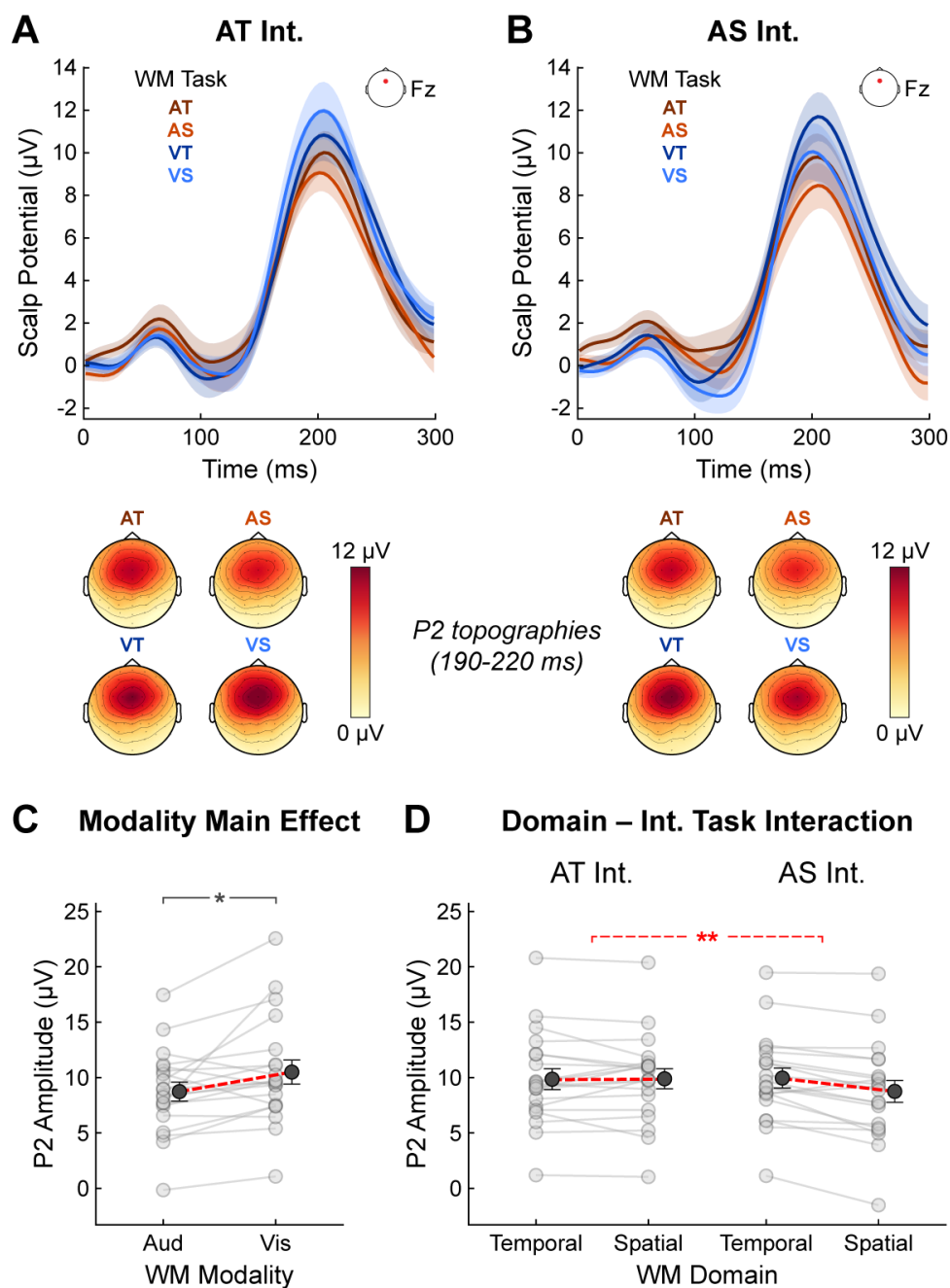
599   below.

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.

600

601

602   Starting with the main effect of WM modality, the amplitudes of P2 components elicited

603   by both auditory Intervening tasks were larger when *visual* information was stored in WM than

604   when the WM task was auditory (Fig. 5C). This could reflect a greater availability of auditory

605   processing resources to respond to the Intervening task stimuli when these resources had not

606   previously been taxed by an auditory WM task. An alternative explanation is that the auditory

607   stimuli in the WM task caused habituation of the auditory Intervening task ERPs, resulting in

608   suppressed P2 amplitudes. While the 2-second gap between WM encoding offset and Intervening

609   task onset is close enough to reduce ERP amplitudes (Budd et al., 1998), such habituation would

610   likely have been reset by the acoustic differences between the WM and Intervening task stimuli,

611   similar to what occurs during auditory oddball paradigms (Paavilainen, 2013). In addition, ERPs

612   elicited by the *final* AT Intervening task stimuli were modulated in a similar fashion to the onset

613   ERPs (see Supplemental Fig. 1), which is inconsistent with a simple habituation account. A similar

614   modulation of the final Intervening task ERPs was not observed in the AS task, but in this task,

615   participants could sometimes make their decision before the final stimulus by comparing the first

616   and second stimulus locations. Thus, the observed modulation of P2 amplitudes likely reflects

617   resource availability for Intervening task processing, rather than auditory response habituation.

618   P2 amplitudes were also affected by domain similarity between the tasks, as evidenced by

619   the significant interaction between WM domain and Intervening task. Specifically, P2s elicited by

620   the first AS Intervening task stimulus were suppressed when the WM task was also spatial; this

621   effect was, on average, smaller for the AT Intervening task (Fig. 5D). Although the three-way

622   interaction involving WM task modality did not reach statistical significance, the two-way

623   interaction appears to have been driven by the visual WM conditions, in which P2s were larger

624   when the domain differed between the two tasks. To demonstrate this, we split the P2 peak data

625   into auditory and visual WM conditions, then conducted separate 2-way ANOVAs with factors of

626   WM domain and Intervening task. The significant interaction between these terms remained in the

627   visual WM conditions (F = 10.77, df = 19, p = 0.004; beneath a Bonferroni-Holm adjusted alpha

628   criterion to account for the two additional ANOVAs, assuming an initial alpha criterion of p =

629   0.05). In the auditory WM conditions, no significant main effects or interactions were detected

630   (even when uncorrected for multiple comparisons). In sum, P2 components of the ERPs scaled

631   inversely with dual-task similarity in modality and domain. P2s amplitudes were generally higher

632   when the tasks were mismatched on either dimension, leading to the largest P2s when the tasks

633   differed in both modality and domain.

634

635   *Alpha-band oscillatory activity*

636   During WM encoding, oscillatory power in the alpha (8-13 Hz) frequency band steadily

637   increased relative to baseline in the auditory WM conditions, and decreased relative to baseline in

638   the visual WM conditions (Supplemental Fig. 2). These patterns of alpha power persisted into the

639   WM retention window, shown in Fig. 6. Preliminary analysis of alpha power in the retention

640   window using cluster-based permutation testing revealed no significant differences based on the

641   domain of the WM task, so data were collapsed across this dimension for the following analyses.

642   The grand average power spectrum during the pre-trial baseline, showing a clear peak in the alpha

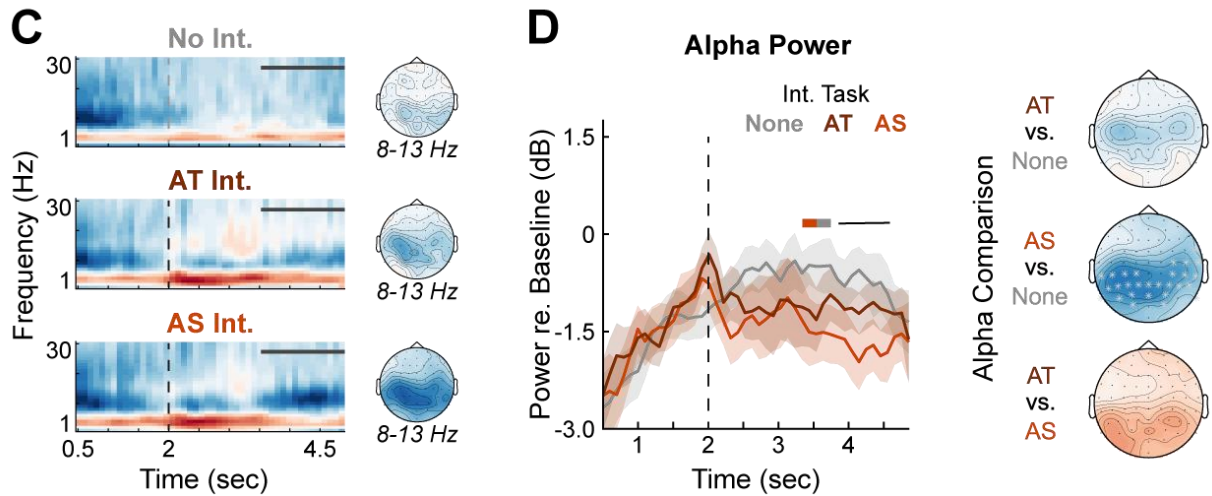643   band at a parietal electrode site, can be found in Supplemental Fig. 3.

644
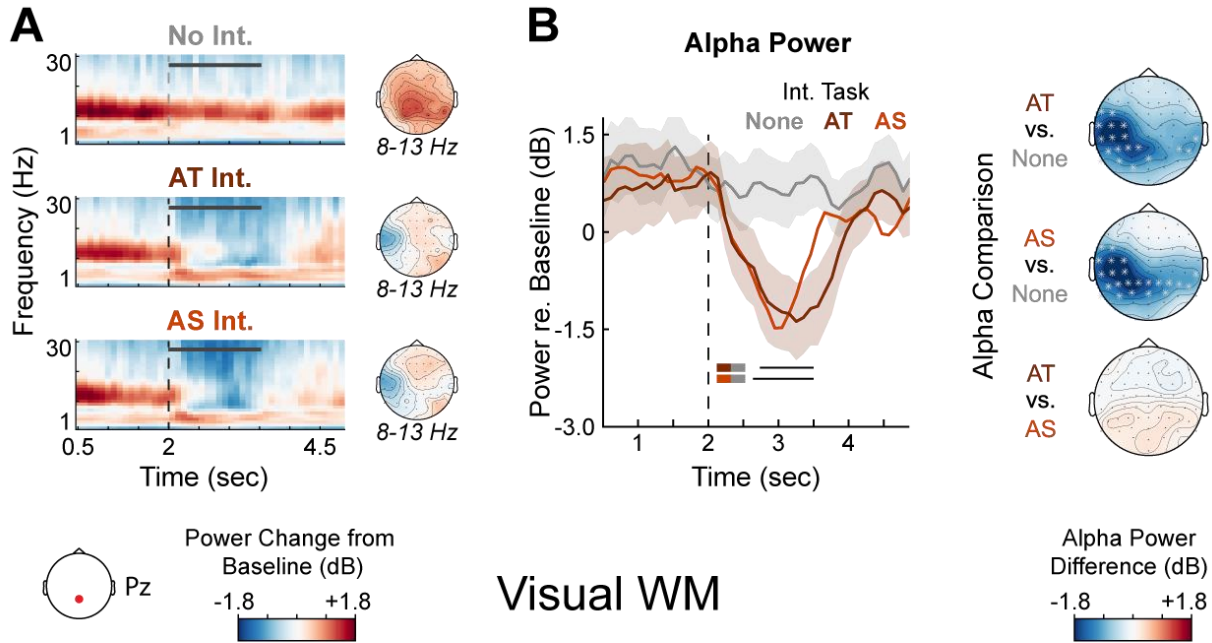
Figure 6: Alpha power results. Grand-average time-frequency responses are shown at a parietal electrode site (Pz) during the memory retention window for each of the three Intervening tasks and for auditory (**A**) and visual (**C**) WM. Responses are averaged across WM domain and shown as dB change relative 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

646 represent S.E.M. Black horizontal bars indicate the temporal extent of significant clusters for the
647 comparisons shown in the topoplots to the right. When significant clusters were detected, the alpha power
648 difference was plotted only in the significant time region; otherwise, the topoplot time regions span the
649 entire cluster test window. * indicates channel membership in a significant cluster (p < 0.05).
650
651
652   In the auditory WM tasks, alpha power remained elevated relative to baseline throughout

653 the memory retention window when no Intervening task was present (Fig. 6A, top panel).

654 However, the onset of both the AT and AS Intervening tasks suppressed these ongoing alpha

655 oscillations (middle and bottom panels). We examined this effect by conducting cluster-based

656 permutation tests on the alpha power time courses (see Materials and Methods) in two time regions

657 of interest: one spanning the onset of the Intervening task through the end of the Intervening task

658 response window (2 to 3.5 seconds into WM retention), and another from 3.5 to 5 sec, during

659 which participants needed to start retrieving the information encoded in WM in anticipation of the

660 probe sequence. During both of the Intervening tasks (in the earlier window), alpha power was

661 significantly reduced relative to the no Intervening task condition (AT Intervening task, p = 0.005;

662 AS Intervening task, p = 0.003; Fig 6B). With the statistical threshold for generating the initial

663 clusters set at p = 0.01, no differences were found between alpha power in the two Intervening

664 tasks. However, relaxing this threshold to p = 0.05 revealed a late AT vs. AS Intervening task

665 cluster between 3.14 and 3.45 sec (p = 0.026; not shown). This indicates that the alpha power

666 reduction lasted longer when the Intervening task was AT than when it was AS, probably because

667 participants could make their perceptual judgment earlier in the AS Intervening task than the AT

668 Intervening task. In the auditory WM conditions, no significant clusters were found in the later

669 (3.5-5 sec) time window.

670   In the visual WM conditions without an Intervening task, alpha power gradually returned

671 from the suppression during WM encoding to near-baseline levels during WM retention (Fig. 6C,

672 top panel). On trials with an AS Intervening task, there was a tendency for alpha to become

673   suppressed again after the conclusion of the Intervening task (bottom panel). We examined these

674   alpha power time courses using cluster-based permutation testing restricted to the same time

675   windows as in the auditory WM conditions: 2 to 3.5 sec and 3.5 to 5 sec from the start of the WM

676   retention phase. Unlike the auditory conditions, no significant differences were observed during

677   the Intervening tasks. However, in the later window, when the Intervening task was AS, alpha

678   power was significantly reduced relative to when there was no Intervening task ($p = 0.011$; Fig.

679   6D). This alpha reduction did not reach significance following the AT Intervening task. When the

680   cluster formation threshold was again relaxed to $p < 0.05$, alpha power was found to be

681   significantly lower following the AS Intervening task than the AT task ($p = 0.027$, not shown). In

682   other words, the alpha suppression present during visual WM encoding resumed following the

683   spatial Intervening task, possibly representing a signature of domain-driven interference.

684   Finally, the presence of either Intervening task also caused an increase in frontal theta (4-

685   7 Hz) power throughout the memory retention window (Supplemental Fig. 4). In all but one of the

686   conditions with an Intervening task, cluster-based permutation tests revealed at least one time-

687   channel cluster in which theta power was significantly elevated relative to the corresponding no

688   Intervening task conditions ($p < 0.05$ for all clusters). In the remaining condition (AT WM, AS

689   Intervening task), a similar increase in theta power was trending toward significance ($p = 0.053$).

690   Although ERPs can manifest as activity in this frequency range as well, the increase in theta power

691   was observed both preceding and well after the Intervening task, meaning it cannot easily be

692   explained by evoked responses. Frontal-midline theta power and theta-band phase coupling in

693   fronto-parietal networks have been linked to task-switching (Cunillera et al., 2012; López et al.,

694   2019; Sauseng et al., 2006) and high levels of cognitive load (Gevins, 1997; McEvoy et al., 2000),

695   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-

719   engage WM mechanisms more strongly after interruption of visual WM rehearsal by spatial

720   processing in a different sensory modality.

721

722   *Similarity in sensory modality leads to increased dual-task interference*

723   While this is the first study we are aware of to investigate how both sensory modality and

724   information domain influence dual-task interference, several previous behavioral studies have

725   determined that interference depends on some axis of similarity between two simultaneous tasks.

726   For instance, Kim et al. (2005) reported increased Stroop task interference when WM was loaded

727   with verbal information that overlapped with the Stroop target. On the other hand, there was no

728   effect of loading WM with spatial information irrelevant to the Stroop task, and loading WM with

729   information similar to the Stroop task distractor actually reduced interference. Similarly, when

730   participants have to retrieve both the orientation and location of visual items, visual distractors that

731   differ from the target only in orientation selectively impair retrieval of target orientation, but not

732   location (Marini et al., 2017). In the realm of phonological WM, distractor words presented during

733   WM retention selectively interfere with recall of words with which they share phonological

734   properties (Oberauer & Lange, 2008).

735   In line with the current study, previous work has specifically implicated similarity in

736   sensory modality as a determinant of distractor and dual-task interference. For instance, auditory

737   and visual discrimination thresholds (for pitch and contrast, respectively) are unaffected by

738   concurrent, irrelevant stimuli in the opposite modality, but are considerably higher in the presence

739   of concurrent stimuli in the same modality (Alais et al., 2006). In dual-task paradigms, increased

740   interference has been observed when both tasks are tactile, auditory, or visual, as compared to

741   conditions in which the tasks were presented in different modalities (Morrison et al., 2015; Scerra

742   & Brill, 2012). These results lend support to multiple resource theory, which states that

743   interference between two tasks depends on whether they compete for the same set of limited neural

744   resources.

745

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

747   The neural underpinnings of dual-task interference have been the topic of substantial debate. An

748   opposing view to the multiple resource theory posits that a general, "central bottleneck" limits

749   performance independently of the specific processing demands of the component tasks (Pashler,

750   1998); a number of neuroimaging studies have provided support for this view (Dux et al., 2006;

751   Marois & Ivanoff, 2005; Tombu et al., 2011). However, contrasting studies more in line with

752   multiple resource theory have shown that interference is marked by increased activity in the brain

753   regions involved in processing the component tasks, without recruitment of any additional areas

754   in dual-task conditions (Adcock et al., 2000; Bunge et al., 2000). This has also been observed at

755   the level of single neurons, with dual-task interference being linked to reduced representational

756   fidelity in a shared population of lateral prefrontal cortex neurons required for both tasks

757   (Watanabe & Funahashi, 2014).

758   In the current study, we attempted to differentially load auditory- and visual-biased

759   cortical networks with our dual-task paradigm. Recruitment of these two networks for attention

760   and WM processing is clearly differentiable based on both sensory modality and information

761   domain (Michalka et al., 2015, 2016; Noyce et al., 2017), supporting the view that they are

762   physiologically distinct neural resources. On the other hand, each network is recruited by a variety

763   of tasks within its preferred modality and domain. Thus, two related but unidentical tasks could

764   interfere with each other if they both load onto the same cortical network, consistent with a "central

765   bottleneck." The coexistence of these properties suggests that the central bottleneck and multiple

766   resource accounts of dual-task interference may not be entirely dichotomous, an idea supported by

767   the present results. Interference patterns measured with behavior and pupillometry were generally

768   similar between the AT and AS Intervening tasks, indicating that both Intervening tasks loaded

769   predominantly onto the same network (likely the auditory-temporal network), despite one

770   involving temporal processing and the other spatial processing (i.e., an apparent central

771   bottleneck). In the WM task on the other hand, dual-task interference patterns were strikingly

772   different between the AT and AS WM conditions. The AS WM task required participants to map

773   absolute sound locations into exocentric coordinates, which likely drove the visual-spatial network

774   more than the AS Intervening task, which only required an immediate comparison of two sets of

775   interaural cues. In other words, the AS WM task was likely able to recruit a separate pool of neural

776   resources, reducing interference from the auditory Intervening tasks; this result is more consistent

777   with multiple resource theory. Thus, there may not be a clear and generalizable resolution between

778   these two theories; rather, dual-task interference effects likely depend on relationships between the

779   specific tasks used, and the specific brain regions and networks these tasks recruit.

780   Another noteworthy aspect of the results presented here is that the processing costs of

781   loading onto a single network (measured as increased behavioral interference, larger pupil

782   diameter, and ERP amplitude suppression) were greater than the costs of switching between

783   networks. This outcome is not entirely obvious, as task-switching also incurs a processing cost

784   (Arnell & Jolicœur, 1999; Chun & Potter, 2001; Hsieh & Allport, 1994; Meiran et al., 2000). Our

785   auditory Intervening tasks did modestly impair visual WM retrieval, consistent with a behavioral

786   cost of switching between the complementary networks. However, this effect was relatively weak

787   compared to the behavioral (and physiological) costs when the tasks loaded onto shared network

788   resources. Thus, the ability to simultaneously leverage complementary attention and WM networks

789   appears to confer a processing benefit, which may be especially relevant in complex, multisensory,

790   real-world environments.

791

792   *Separate contributions of behavioral, pupillometry, and EEG measures*

793   Measures of behavior, pupillometry, and EEG in this study each offered unique insights into the

794   dual-task interference at play. Pupillometry was included because it can serve as a sensitive

795   biomarker of the cognitive effort required to achieve a certain performance level, even in the

796   absence of behavioral differences. For instance, maximum pupil dilations evoked by speech are

797   larger for listeners with hearing loss than normal-hearing controls, even when the signal-to-noise

798   ratio is high enough for both groups to achieve ceiling performance. This is thought to reflect a

799   heightened degree of listening effort, not captured by behavioral measures alone (Ohlenforst et al.,

800   2017). Pupil dilations have also been shown to correlate with participants' subjective assessments

801   of task difficulty (Koelewijn et al., 2015). In the present study, this benefit manifested in our ability

802   to detect differences between the AT and AS WM conditions when the Intervening task was AS.

803   This Intervening task was easy enough that a majority of participants did not make errors on it in

804   any WM condition. However, substantial differences in the amplitude of pupil dilations elicited

805   by the AS Intervening task indicated that the information in WM did cause processing interference,

806   which participants overcame to achieve excellent performance.

807   In addition, behavioral performance on each trial is often a single-sample snapshot,

808   encompassing all the processing dynamics present in the task. Pupillometry allows for more fine-

809   grained dissection of the time course of these dynamics. For instance, pupil data in this study

810   revealed that participants were modulating their effort level to encode AT information in WM

811   when an auditory Intervening task would be present, indicating preparation for a challenging

812   interference condition. Analysis of the temporal dynamics of pupil responses is becoming more

813   common in other fields as well, including the study of cognitive processes that unfold through time

814   like decision making (Satterthwaite et al., 2007) and the perception of temporally complex inputs

815   like speech and music (Jagiello et al., 2019; Winn & Moore, 2018). Conversely, though, the

816   behavioral data provided insights that would have been difficult to glean from pupillometry alone.

817   Pupil responses elicited by the Intervening task stimuli were drastically different from the no

818   Intervening task conditions, making it difficult to interpret the magnitude of the interference

819   effects. The behavioral data puts this in more concrete terms; we could measure exactly how much

820   performance on each task was influenced by the other. Because of these complementary benefits,

821   we argue that pupillometry data are best interpreted alongside corresponding behavioral measures.

822   In the visual WM conditions, interference by the auditory Intervening tasks was generally

823   low, and neither behavioral nor pupil responses differed based on whether the information held in

824   WM was temporal or spatial. This finding was counter to our expectations; for instance, we

825   expected to observe mild interference when the Intervening and WM tasks were both spatial.

826   Auditory and visual spatial processing are known to be represented in the EEG response, often in

827   the form of lateralized alpha oscillations at parietal scalp locations (Bednar & Lalor, 2018; Deng

828   et al., 2020; Kelly et al., 2006; Klatt et al., 2018; Thut et al., 2006; Tune et al., 2018; Worden et

829   al., 2000). Given the sensitivity of EEG to spatial processing, we expected to observe domain-

830   based interference patterns in our EEG data when the WM and Intervening tasks were both spatial.

831   Indeed, ERPs and alpha modulations proved particularly sensitive to these subtle interference

832   patterns, which did not affect error rates or pupil responses. The P2 components of ERPs were

833   largest when the tasks were presented in different modalities and, in the case of the visual WM

834   conditions, when the information domains of the tasks were also mismatched. This pattern is

835   consistent with greater availability of processing resources when the tasks did not compete for

836   shared modality- and domain-biased neural resources (Kok, 1997).

837   We found that alpha power was modulated by the presence of an Intervening task and by

838   its information domain. This alpha had a more anterior and centralized distribution that the

839   lateralized alpha often implicated in spatial attention (Fig. 6). Alpha power with the scalp

840   distribution we found in this study has been linked to various types of WM load, including visually

841   presented letters and numbers (Jensen et al., 2002; Klimesch et al., 1999; Schack & Klimesch,

842   2002), shape information (Herrmann et al., 2004; Johnson et al., 2011), spatial information

843   (Bastiaansen et al., 2002), and auditory stimuli (Lim et al., 2015; Obleser et al., 2012; Vogt et al.,

844   1998). In the auditory WM conditions, the auditory Intervening tasks interrupted this alpha activity

845   during WM retention. This is consistent with previous work showing that distracting visual stimuli

846   interrupt alpha power during retention of both auditory and visual information in WM (Hakim et

847   al., 2020; Mishra et al., 2013). In the visual WM conditions, alpha power was suppressed relative

848   to baseline during WM encoding, perhaps due to the need to attend task-relevant visual stimuli

849   (Foxe & Snyder, 2011; Jensen & Mazaheri, 2010). In these conditions, there was no significant

850   change in alpha power during Intervening task presentation; however, following the spatial

851   Intervening task, the alpha suppression resumed. The functional implications of these alpha

852   suppressive effects, and why their pattern differed between the auditory and visual WM conditions,

853   are topics warranting targeted future investigation.

854

855   **Conclusions and future directions**

856   In a dual-task paradigm designed to drive auditory- and visual-biased attention and WM networks,

857   we reported behavioral, autonomic, and electrophysiological signatures of interference when the

858   tasks drew upon shared neural control resources. Future studies should implement a similar

859   paradigm with fMRI to pinpoint the neural networks involved in this kind of dual-task interference,

860   as well as invert the paradigm used here – using visual rather than auditory Intervening tasks – to

861   test whether interference patterns flip under opposite network demands.

862

869

870

871

872

873

874

875

876

877

878

879

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

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

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-

919   9780198505150

920   Cunillera, T., Fuentemilla, L., Periañez, J., Marco-Pallarès, J., Krämer, U. M., Càmara, 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-

937   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

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

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

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

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-

1062   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

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 ofhuman 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**
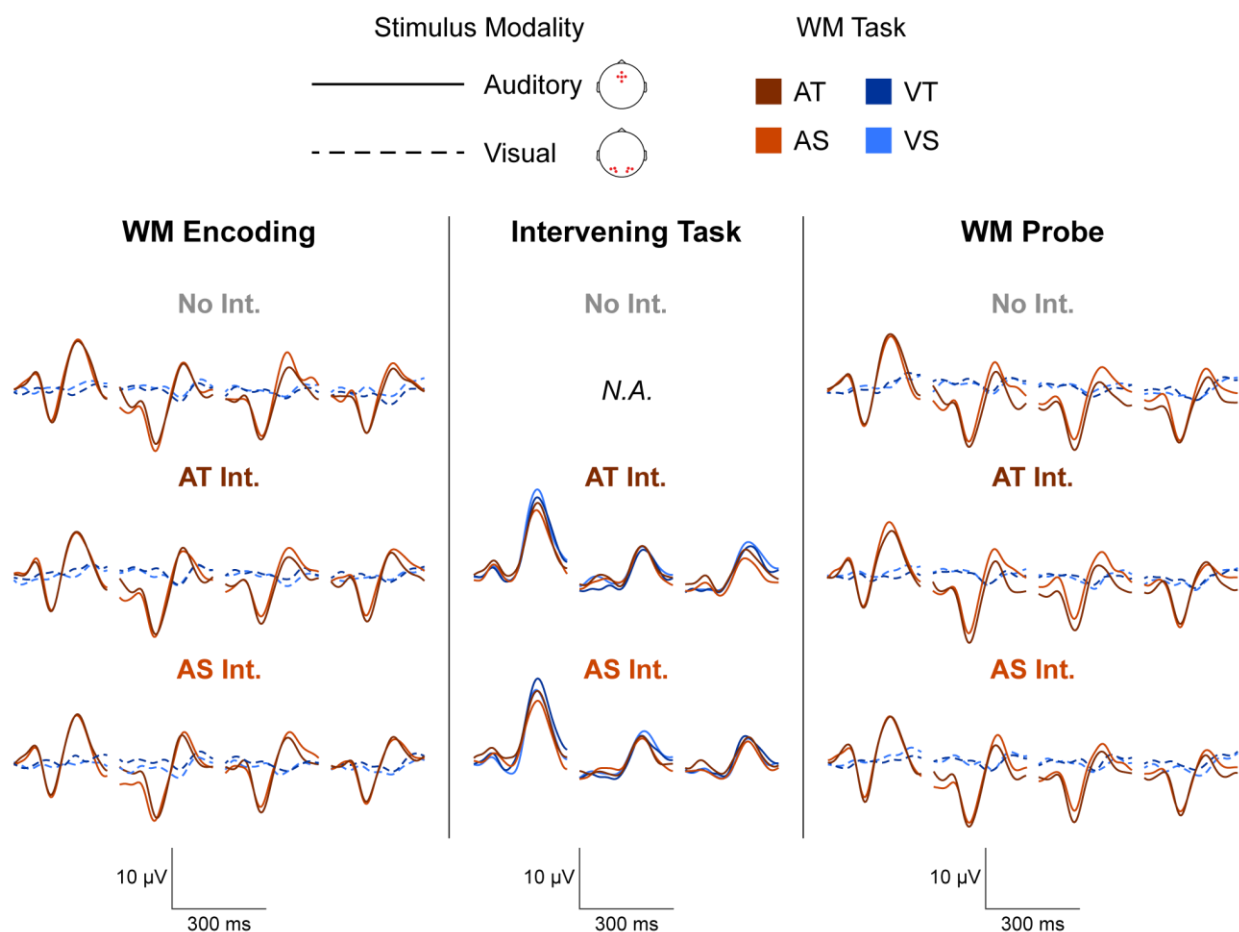
1164

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
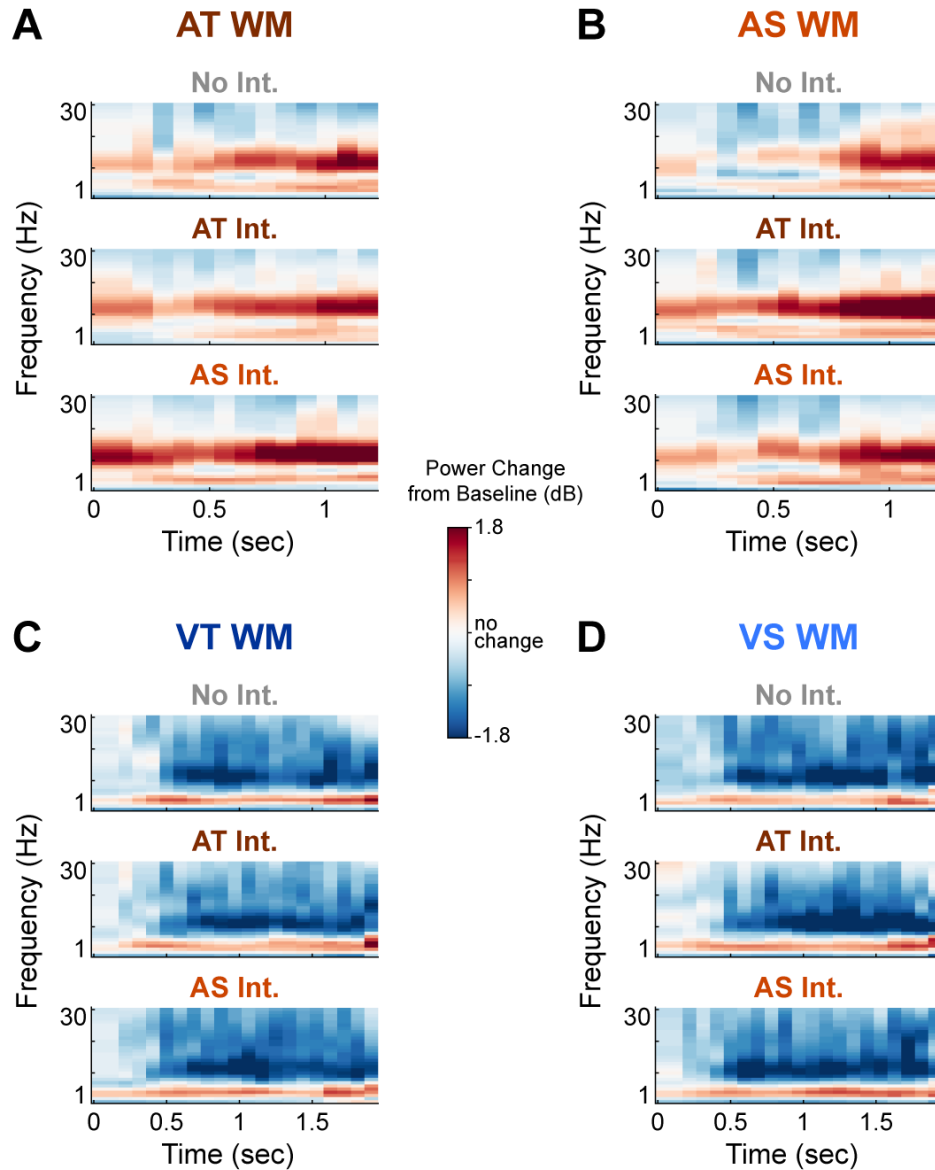
1172

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.

1180
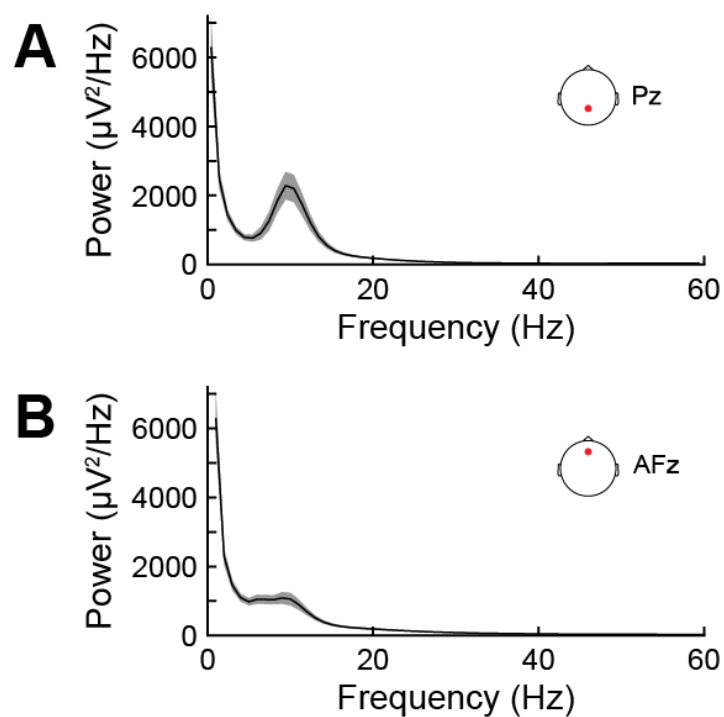
Supplemental Figure 3: Grand average power spectra during the pre-trial baseline, plotted separately at channel Pz (**A**) and channel AFz (**B**). Note the well-defined peak in the alpha (8-13 Hz) frequency range at the parietal electrode site. Theta (4-7 Hz) power was more task-induced and so was less 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
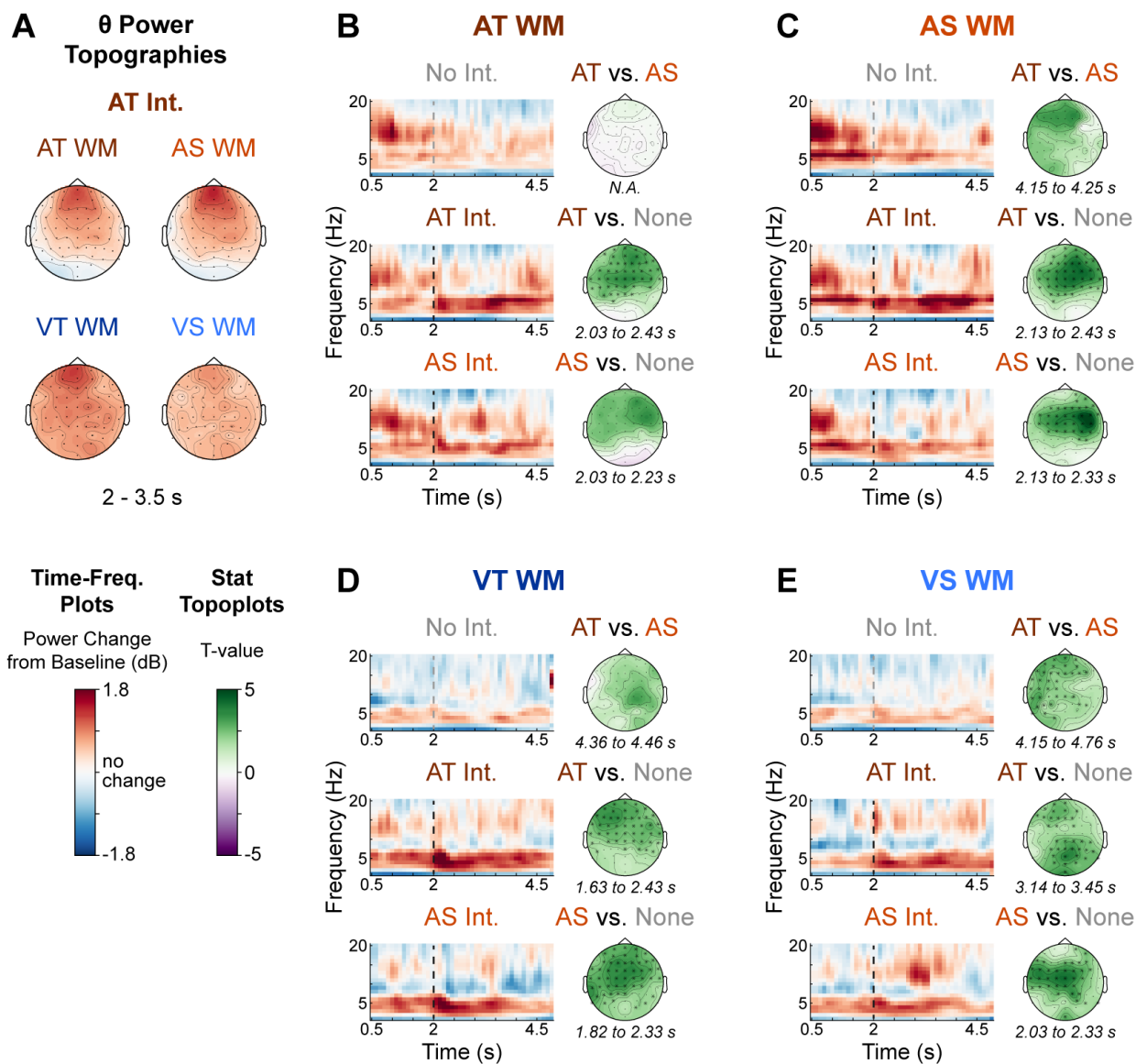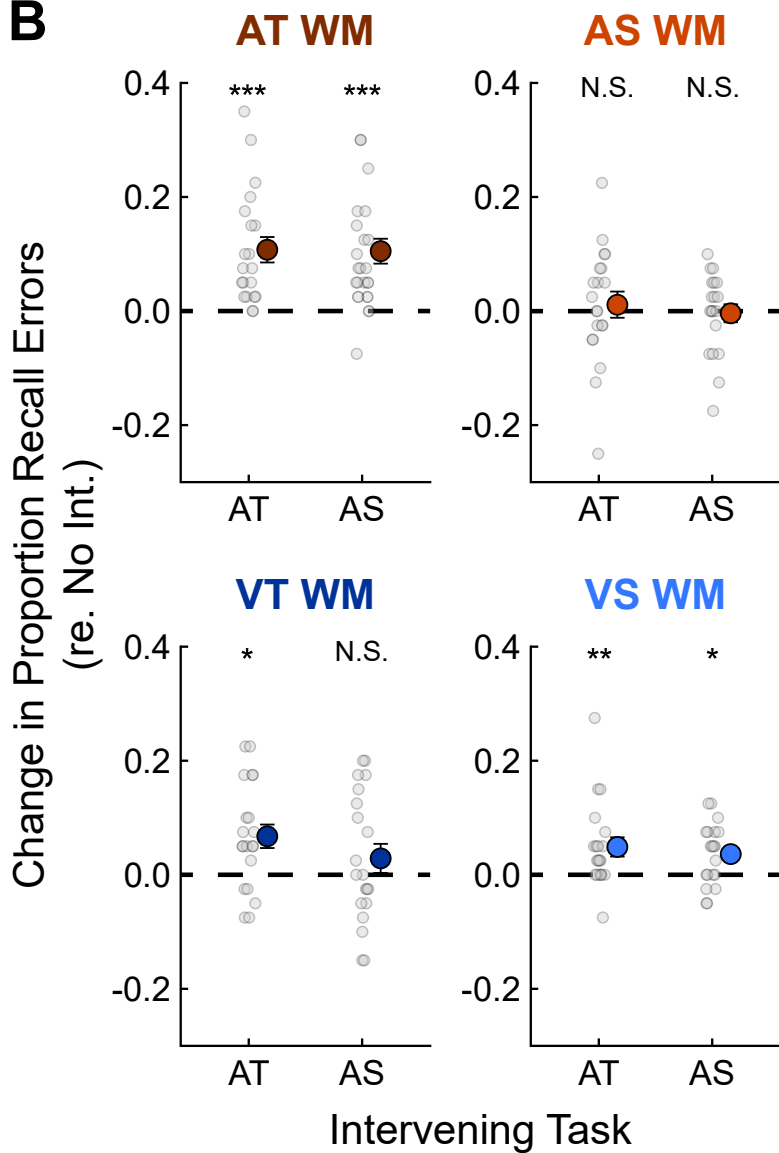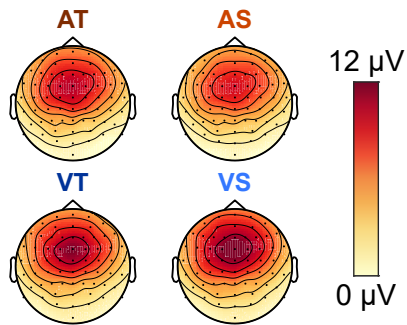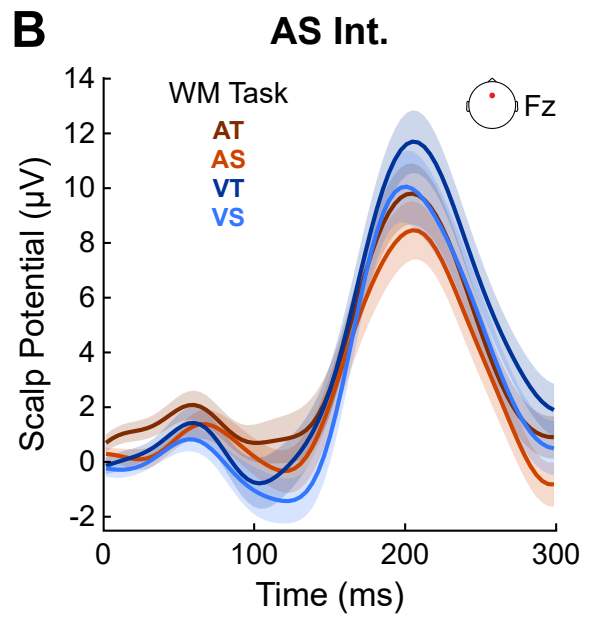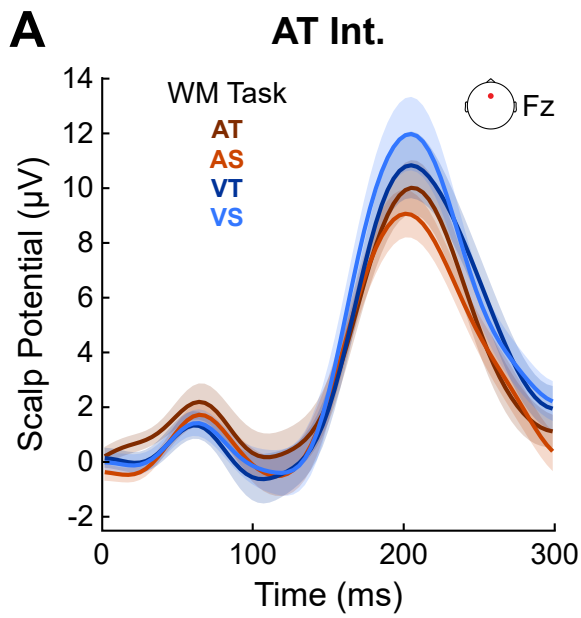
Supplemental Figure 4: Increased theta (4-7 Hz) power in conditions with an Intervening task. A, Topographies of theta power when the Intervening task was AT, shown to illustrate the fronto-central concentration of theta power. **B-E,** Left panels show time-frequency responses at a fronto-central electrode site (AFz) relative to the pre-trial baseline. Within each WM task, scalp topographies to the right show the results of cluster-based permutation tests comparing the indicated Intervening task conditions. T-values from cluster formation are plotted on the green-purple axis, as in some cases, outlier participants with Supplemental Figure 3.12 (Continued): anomalously strong theta power changes exaggerated the apparent differences in the grand average time-frequency plots. Asterisks signify membership in a significant cluster (p < 0.05). For each comparison, only the strongest time-channel cluster is shown, although some comparisons did yield multiple significant clusters. If the strongest 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
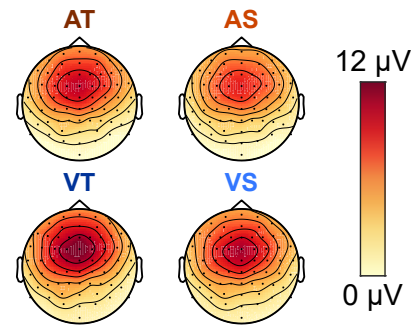
1209   (time shown as N.A.).

A

-4°   0°
-45°   45°
-90°   90°
1.5 m

B

WM Encoding → Hold in memory... → WM Probe
WM Encoding → Intervening Task → WM Probe

subject responses

0   2   ~3   5.5
Time (sec)

C

Working Memory Task

Encoding

Auditory *or* Visual

Probe

Auditory *or* Visual

Time

Time

Encoding vs. Probe
Temporal: "Different"
Spatial: "Same"

D

Intervening Task

-4° 0°

A   B   C

⊢1⊣   ⊢ 2 ⊣

Time

Auditory-Temporal
Interval 1 or 2 longer?
"2"

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

A

No Intervening Task

B

AT WM   AS WM

VT WM   VS WM

Change in Proportion Recall Errors
(re. No Int.)

Intervening Task

A

No Int.

AT Int.

AS Int.

Pupil Diameter (z)

Time re. Start of WM Retention (sec)

p < 0.05

Intervening task onset

WM Task

AT   VT
AS   VS

Difference Plots

Intervening Task

minus

No Intervening Task

B   **Differential effect of Int. tasks**

Int. Task: *AT − None*   Int. Task: *AS − None*

Pupil Diameter Difference (z)

p < 0.05

Time re. Start of WM Retention (sec)

A AT Int.

WM Task
AT
AS
VT
VS

Fz

Scalp Potential (μV)

Time (ms)

B AS Int.

WM Task
AT
AS
VT
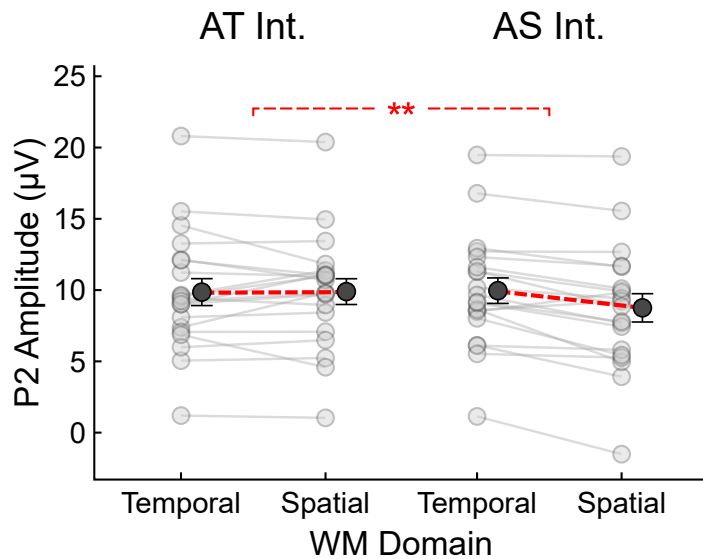VS

Fz

Scalp Potential (μV)

Time (ms)

AT   AS
VT   VS

12 μV

0 μV

AT   AS
VT   VS

12 μV

0 μV

P2 topographies (190-220 ms)

C Modality Main Effect

P2 Amplitude (μV)

*

Aud   Vis
WM Modality
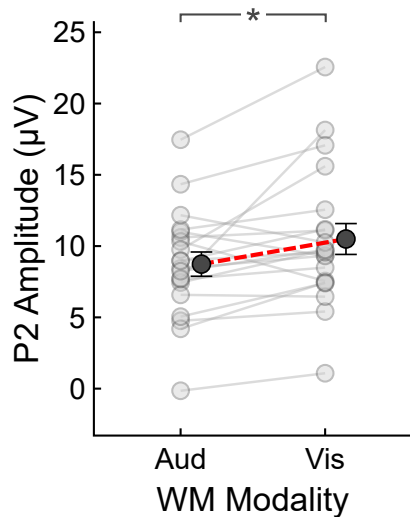
D Domain – Int. Task Interaction

AT Int.   AS Int.

P2 Amplitude (μV)

**

Temporal Spatial Temporal Spatial
WM Domain

Auditory WM

A

No Int.

8-13 Hz

AT Int.

8-13 Hz

AS Int.

8-13 Hz

B

Alpha Power

Int. Task
None   AT   AS

Alpha Comparison

AT
vs.
None

AS
vs.
None

AT
vs.
AS

Pz

Power Change from
Baseline (dB)

-1.8   +1.8

Visual WM

Alpha Power
Difference (dB)

-1.8   +1.8

C

No Int.

8-13 Hz

AT Int.

8-13 Hz

AS Int.

8-13 Hz

D

Alpha Power

Int. Task
None   AT   AS

Alpha Comparison

AT
vs.
None

AS
vs.
None

AT
vs.
AS