Behavior Research Methods (2024) 56:5959–5985
https://doi.org/10.3758/s13428-023-02328-6

ORIGINAL MANUSCRIPT

Modality matters: Three auditory conflict tasks to measure individual differences in attention control

Alexander P. Burgoyne[1] · **Dolly T. Seeburger**[1] · **Randall W. Engle**[1]

Accepted: 22 December 2023 / Published online: 16 February 2024

Abstract

Early work on selective attention used auditory-based tasks, such as dichotic listening, to shed light on capacity limitations and individual differences in these limitations. Today, there is great interest in individual differences in attentional abilities, but the field has shifted towards visual-modality tasks. Furthermore, most conflict-based tests of attention control lack reliability due to low signal-to-noise ratios and the use of difference scores. Critically, it is unclear to what extent attention control generalizes across sensory modalities, and without reliable auditory-based tests, an answer to this question will remain elusive. To this end, we developed three auditory-based tests of attention control that use an adaptive response deadline (DL) to account for speed–accuracy trade-offs: Auditory Simon DL, Auditory Flanker DL, and Auditory Stroop DL. In a large sample ($N = 316$), we investigated the psychometric properties of the three auditory conflict tasks, tested whether attention control is better modeled as a unitary factor or modality-specific factors, and estimated the extent to which unique variance in modality-specific factors contributed incrementally to the prediction of dichotic listening and multitasking performance. Our analyses indicated that the auditory conflict tasks have strong psychometric properties and demonstrate convergent validity with visual tests of attention control. Auditory and visual attention control factors were highly correlated ($r = .81$)—even after controlling for perceptual processing speed ($r = .75$). Modality-specific attention control factors accounted for unique variance in modality-matched criterion measures, but the majority of the explained variance was modality-general. The results suggest an interplay between modality-general attention control and modality-specific processing.

Keywords Attention control · Executive functions · Sensory modalities · Individual differences · Reliability paradox

> *"Every one knows what attention is…It implies withdrawal from some things in order to deal effectively with others"* (James, 1890, p. 404).

Selective attention has fascinated psychologists for well over a century, but scientific interest in the construct was ignited by the cognitive revolution of the 1950s (Sperry, 1993). Influential early research used an auditory-based paradigm called *dichotic listening* to better understand the nature of capacity constraints on cognitive performance (Cherry, 1953). In a dichotic listening test, subjects are presented with different auditory streams to each ear and asked to repeat aloud (i.e., "shadow") the message delivered to one ear while ignoring the other. Subjects are generally unaware of content presented to the unattended ear, but this is not always the case: when the test taker's name is presented to the unattended channel, some subjects notice it (Cherry, 1953; Conway et al., 2001; Wood & Cowan, 1995). Subjects who notice their own name in the unattended channel typically demonstrate worse shadowing performance immediately following its presentation (Conway et al., 2001; Wood & Cowan, 1995), suggesting that attentional resources are "captured" by the distractor and allocated to the to-be-ignored channel for a short period of time. Furthermore, subjects with lower cognitive ability—and in particular, those with lower working memory capacity—are more likely to demonstrate this attentional capture effect (Conway et al., 2001). That is, lower-ability subjects display more failures of selective attention due to highly salient but task-irrelevant distractors.

The dichotic listening paradigm established two major findings. First, it provided evidence for *attenuation theory*, the idea that unattended information is still processed to some degree even if it does not always reach conscious awareness (Anderson, 2005; Treisman, 1964). Second, it showed that individuals differ in the ability to control

---

✉ Alexander P. Burgoyne
burgoyn4@gmail.com

[1] Georgia Institute of Technology, Atlanta, GA, USA

attention to filter out irrelevant information. Conway et al. (2001) suggested that the critical ability tapped by dichotic listening might be the inhibition of distractors. In particular, they noted the strong theoretical link between inhibition and working memory capacity; individuals with greater working memory capacity might be better able to suppress interference in order to maintain focus on goal-relevant information.

Today, there is strong empirical support for the *executive attention view* of working memory capacity, which argues that the ability to control attention to inhibit distractors is fundamental to cognitive performance (Burgoyne & Engle, 2020; Engle, 2002; Engle, 2018; Gazzaley & Nobre, 2012). According to the executive attention view, the primary reason measures of working memory capacity predict cognitive performance or real-world outcomes (Mashburn et al., 2023) is because they capture individual differences in the ability to control attention (i.e., the "central executive" in Baddeley's (1996) model). If working memory capacity reflects the interplay between attention control and short-term storage, attention control is the primary bottleneck that constrains performance across cognitive domains (Burgoyne et al., 2022). For example, studies have found that the relationships between working memory capacity and fluid intelligence, processing speed, and sensory discrimination ability can be significantly reduced or statistically eliminated by accounting for individual differences in attention control (Burgoyne et al., 2022; Burgoyne et al., 2023; Draheim et al., 2021; Tsukahara et al., 2020). This suggests that attentional abilities may underpin the *positive manifold* (Spearman, 1904)—the positive correlations observed among different broad cognitive ability factors.

Given the theoretical importance of attention control, a plethora of tasks have been developed to measure individual differences in this ability (Burgoyne et al., 2023; Draheim et al., 2021). However, in contrast to foundational work in this area (e.g., Cherry, 1953), the vast majority rely on visual stimuli. For example, in the antisaccade task (Hallett, 1978), subjects must ignore a flickering asterisk on one side of a computer screen and rapidly shift their focus to the opposite side of the screen to identify a briefly presented letter. In the selective visual arrays test (Vogel et al., 2005), subjects must attend to a subset of items in a visual display (e.g., remember red items, ignore blue items) to determine whether those items change in a subsequent visual display. Three "conflict tasks"—the Stroop, Flanker, and Simon paradigms (Eriksen & Eriksen, 1974; Simon & Rudell, 1967; Stroop, 1935)—are also frequently used to measure individual differences in attention control (Burgoyne et al., 2023). In the Stroop task, subjects must name the color that a word is printed in while ignoring its meaning. In the Simon task, they must determine which direction an arrow is pointing while ignoring what side of the computer screen it appears on. In the Flanker task, subjects must determine which direction a central arrow is pointing while ignoring the direction of the flanking arrows.

All of these tasks demand selective attention and inhibition of irrelevant stimulus characteristics. However, because they all use visual stimuli, it is unclear whether the executive attention view also extends to the auditory modality, or other sensory modalities. More broadly, it is unclear to what extent attention control is modality-general or modality-specific, and without reliable auditory-based tests, an answer to this question will remain elusive.

The question of whether attention control is modality-general or modality-specific echoes investigations into the domain-generality of working memory capacity in the 1990s and early 2000s. For example, Shah and Miyake (1996) appeared to show a dissociation between spatial and verbal working memory abilities, albeit in small samples of undergraduates at Carnegie Mellon (*N*s = 54 and 60 in two experiments). Specifically, they found that tests of working memory capacity that used verbal stimuli (e.g., reading span) correlated more strongly with verbal outcome measures (e.g., verbal standardized test scores) than with spatial outcome measures, whereas tests of working memory capacity that used spatial stimuli (e.g., rotation span) correlated more strongly with spatial outcome measures than with verbal outcome measures. Shah and Miyake (1996) interpreted these results as evidence for separate pools of resources devoted to language processing and spatial thinking. Furthermore, they suggested that it may be these domain-specific resources that largely account for the predictive validity of working memory measures. This conclusion stands in stark contrast to the executive attention perspective (Engle, 2002), which views domain-general attention as central to performance.

Shortly thereafter, Kane et al. (2004) addressed Shah and Miyake's (Shah & Miyake, 1996) findings. They noted that some of the correlations on which Shah and Miyake (1996) based their conclusions varied greatly in subsequent studies (e.g., Friedman & Miyake, 2000; Sohn & Doane, 2003), casting doubt on the validity of their point estimates and the conclusions that followed from them. They also noted that Shah and Miyake (1996) used a single task to measure verbal-specific or spatial-specific working memory capacity, confounding domain-specificity with the particular task used to measure it. Finally, they echoed a point acknowledged by Shah and Miyake (1996) themselves: By testing a range-restricted sample of undergraduates at a top university, it is possible that differences in domain-general abilities were minimized relative to differences in domain-specific abilities, potentially giving rise to the specific pattern of results they observed.

To address these issues, Kane et al. (2004) had 236 subjects representing a wide range of cognitive ability complete three verbal working memory tests and three spatial working memory tests. They used latent variable modeling to capture variance common to the measures. In a series of confirmatory factor analyses, Kane et al. (2004) found that the best-fitting model had verbal and spatial working memory measures loading on a common factor. Furthermore, when modeled on separate-but-correlated factors, the correlation between working memory factors was very strong ($r = .84$; 95% CI [.80, .87]), indicating that they shared a majority (i.e., 70%) of their variance. Finally, they showed that it was largely variance that was shared across verbal and spatial working memory measures that accounted for the prediction of reasoning performance. Kane et al. (2004) concluded that "the shared variance among measures of WM span and complex cognition reflects primarily the contribution of domain-general attention control, rather than domain-specific storage or rehearsal" (p. 190).

Although Kane et al.'s (2004) conclusion was that the ability to control attention underpins performance across spatial and verbal content areas, it is not necessarily the case that domain-general attention control underpins performance *across visual and auditory sensory modalities*. That said, there is some supporting evidence for cross-modal influences of attention control. For example, Tsukahara et al. (2020) investigated the cognitive underpinnings of sensory discrimination ability. To measure auditory discrimination ability, subjects were presented with two tones, one after another, and asked to make a judgment regarding their relative pitch, duration, or loudness, depending on the task. Tsukahara et al. (2020) found that an attention control factor defined by visual-modality tasks was very highly correlated with auditory discrimination ability (i.e., $r = .79$) and fully mediated its relationship with fluid intelligence. While this work suggests that attention control plays an important role in cross-modal perceptual judgment tasks, it does not necessarily indicate that visual- and auditory-based tests of attention control would tap a common domain-general factor that is predictive of cross-modal outcomes.

Addressing this question could prove challenging. For example, differences in performance on visual or auditory attention control measures could emerge due to differences in early processing of sensory signals, differences in the ability of the attention system to modulate processing across modality-specific sites, or both. As a case in point, a subject with hearing damage might perform better on visual tests, but no conclusion about the nature of attention control as modality-general or modality-specific would be warranted from this evidence. It could be that performance on different-modality tests is supported by both a domain-general attentional system and influences from sensory-specific processing sites. As we will explain, the field needs more reliable tests of attention control combined with experimental approaches that are suited to shed light on this question. Although neuroscientific evidence indicates that a common attentional network may underpin modulation of activity in early visual and early auditory cortex (Dosenbach et al., 2006; Fan, 2014; Green et al., 2011; Spagna et al., 2015; Wu et al., 2020), improving the measurement of attentional abilities in the auditory modality will only enhance this research.

Unfortunately, measuring individual differences in attention control has posed a challenge, even in the visual modality. As is now well known, conflict tasks such as the Stroop paradigm suffer from psychometric problems when used as-is for the study of individual differences in attention control (Hedge et al., 2018). Part of the difficulty is accounting for speed–accuracy trade-offs (Heitz, 2014): Because subjects can differentially prioritize speed versus accuracy, an outcome measure from these tasks should consider both aspects of performance. Traditionally, however, the outcome measure simply subtracts response times on correct congruent trials from response times on correct incongruent trials, overlooking the distributions of incorrect responses (Rouder & Haaf, 2019). Another part of the problem is the use of the aforementioned difference score: subtracting performance in one condition from another leads to an unreliable outcome measure, particularly when performance in the two conditions is highly correlated (Draheim et al., 2019; Overall & Woodward, 1975). Solutions have been proposed that range from creating efficient, gamified, points-based tasks (i.e., the three-minute "Squared" tests of attention control; Burgoyne et al., 2023) or using more sophisticated statistical approaches (e.g., hierarchical modeling; Rouder et al., 2023), to using an adaptive response deadline, described next.

Adaptive response deadline procedure

In an effort to develop more reliable tests of attention control, our lab recently incorporated an *adaptive response deadline* into two classic conflict paradigms, the Stroop and Flanker tasks (Draheim et al., 2021; Draheim et al., 2023). An illustration of the visual Flanker adaptive deadline task is presented in Fig. 1. Subjects must indicate which direction a central arrow is pointing within a fixed amount of time (e.g., 960 ms), termed the response deadline. In the original version of the task (Draheim et al., 2021), the response deadline changed based on performance across blocks of congruent and incongruent trials, whereas in the revised version (Draheim et al., 2023), the response deadline changed based on performance on incongruent trials on a trial-by-trial basis. Overall, if the subject responds accurately before the response deadline, the deadline becomes shorter, requiring quicker responses (i.e., the task becomes more difficult). If the subject responds incorrectly or too slowly, the deadline becomes longer, allowing slower responses (i.e., the task becomes easier). The adaptive deadline uses a staircase procedure with a 3:1 weighted up-to-down step-size ratio,

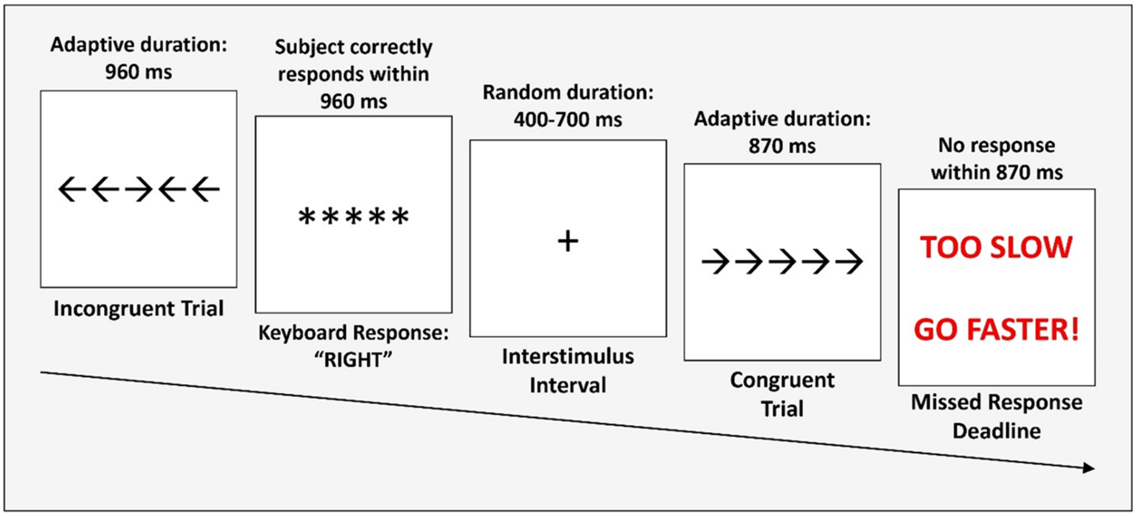

Fig. 1 Example of the adaptive response deadline procedure used in the visual Flanker DL task from Draheim et al. (2021)

and in theory should converge on an accuracy rate of 75% on incongruent trials (Kaernbach, 1991). Thus, accuracy rates are theoretically held constant across participants, midway between ceiling and chance, and the measure of performance—the response deadline at the conclusion of the task—represents the speed at which subjects can maintain the critical accuracy rate. In other words, the measure effectively takes into consideration both the speed and accuracy with which participants can respond, and it is not based on an unreliable difference score. As Draheim et al. (2021) showed, these visual conflict tasks with an adaptive deadline had better psychometric properties than conflict tasks that were scored in the traditional manner, and they also loaded more highly on a common attention control latent factor.

Auditory measures

Although most tests of attention control are based on the visual modality, there are a few auditory attention control measures worth mentioning. One relatively popular paradigm is the Conners Continuous Auditory Test of Attention (Conners, 2014), in which participants must respond when they hear a low tone followed by a high tone, but withhold responses otherwise. Regrettably, the Conners test is proprietary, costly, and used almost exclusively in clinical evaluation (e.g., Wang et al., 2021). We believe researchers would benefit from a set of freely available and reliable tasks in the auditory domain that permit latent variable modeling, thereby allowing researchers to partial out measurement error and draw conclusions closer to the theoretical constructs of interest (Kline, 2023).

To this end, a few research groups have developed auditory analogues of classic conflict tasks, such as an auditory Stroop task (Christensen et al., 2011) and an auditory Flanker task (Chan et al., 2005). We briefly describe these tasks, as the new tools we develop in the present manuscript are based on them. In their auditory Stroop task, Christensen et al. (2011) presented audio messages to participants and had them respond based on either the voice's perceived gender or the gender meaning of the word (e.g., "boy"). They found a significant auditory Stroop effect when subjects had to ignore the voice's gender, but they did not report psychometric statistics, such as the reliability of the measure. This is concerning, because the outcome measure was based on a difference score contrasting congruent and incongruent conditions, which is known to reduce reliability (Overall & Woodward, 1975). In their auditory Flanker task, Chan et al. (2005) presented auditory messages centrally and to the left and right of participants and challenged them to ignore the flanker words ("bat" or "bed") while responding based on the centrally presented word. Once again, no reliability statistics were reported for this measure, although it did demonstrate a significant conflict effect. These studies suggest that conflict effects can emerge in auditory paradigms; however, whether these paradigms would be psychometrically suitable for individual differences research, and in turn, for adjudicating on the nature of attention control as modality-general or modality-specific, is an open question.

The present study

In the present study, we tested the extent to which attention control is modality-general or modality-specific by incorporating an adaptive response deadline into three auditory versions of classic conflict paradigms used to measure attention control: Simon, Flanker, and Stroop. The three auditory

DL (i.e., auditory deadline) tasks are described in detail in the Method section. Using a large sample of participants, we first investigated whether the trio of auditory attention control tests have strong psychometric properties, and then tested whether attention control appears to be modality-general or modality-specific. In doing so, we tested whether the auditory measures of attention control demonstrate convergent validity with visual-based attention control tests above and beyond processing speed, and whether they demonstrate divergent validity with measures of other broad cognitive abilities, including fluid intelligence, working memory capacity, and processing speed. Additionally, we investigated whether variance that distinguishes auditory attention control from visual attention control contributes incrementally to the prediction of two theoretically relevant outcome measures: dichotic listening and multitasking performance. Our interest in multitasking stems from a series of studies showing strong links between attention control and multitasking performance (e.g., Burgoyne et al., 2023; Draheim et al., 2021). We chose dichotic listening and multitasking as our criterion measures because both have been hypothesized to demand controlled attention, but differ in the primary modality of stimulus presentation (i.e., auditory vs. visual). Therefore, we reasoned that they would be an effective foil to estimate the modality-generality or modality-specificity of attention control.

Method

Participants

The study was conducted at the Georgia Institute of Technology in Atlanta, Georgia, USA. All participants were required to be native English speakers and 18–35 years of age. We recruited participants from Georgia Tech, other surrounding colleges in Atlanta, and the broader Atlanta community. Georgia Tech students enrolled in an undergraduate psychology course were given the option to receive 2.5 hours of course credit or monetary compensation for each session. This study was approved by the Georgia Institute of Technology Institutional Review Board under Protocol H20165.

Procedure

Data were collected as part of a larger project, which consisted of more than 40 cognitive tasks administered over five sessions lasting 2.5 hours each. All tasks were computerized, and, with the exception of the multitasks and test of dichotic listening, were programmed using E-Prime (Psychology Software Tools, Pittsburgh, PA). Further information regarding the scope of the data collection effort and other research products based on it can be found at the following link: https://tinyurl.com/ms8htaep.

Participants scheduled each study session according to their own availability, but they were not allowed to complete more than one session on a given day. Participants were paid $200 for completing the five in-laboratory sessions ($30 for session 1, $35 for session 2, $40 for session 3, $45 for session 4, and $50 for session 5). Georgia Tech students were allowed to choose a combination of either financial compensation or research participation credits—the latter is required by some undergraduate psychology courses at Georgia Tech. Participants who frequently rescheduled, missed appointments, or regularly failed to follow directions were not invited back for subsequent sessions.

During data collection, participants were seated in individual testing rooms with a research assistant assigned to proctor each session. The research assistant's job was to run each cognitive test, ensure the participants understood the instructions, and make sure participants were following the rules of the lab, such as not using their phone during the study. The research assistants took extensive notes on participant conduct, which were used to make decisions about data exclusions described below. Up to seven participants could be tested in a given session, although typically two to four participants were scheduled for each time slot.

Demographics

Participants were asked to report their age, gender, and ethnicity. They were asked whether English was the first language they learned and the age at which they learned it, and whether they were fluent in other languages. Participants were asked to report the highest level of education they had achieved as well as their annual household income. Participants were asked whether they had corrected vision, and also whether they had any conditions (e.g., illness, disability, medication use) that might affect their performance on cognitive tasks.

Auditory attention control

Auditory Simon with adaptive response deadline (Auditory Simon DL) In Auditory Simon DL (Fig. 2), subjects must indicate which ear received an auditory stimulus while ignoring the semantic content of the stimulus. Subjects wore headphones for the test. They were instructed to press the "P" key to indicate that the auditory stimulus was delivered to the right ear and the "Q" key to indicate that the stimulus was delivered to the left ear. These instructions were displayed on the computer screen for the duration of the task. The auditory stimulus was the word "left" or "right" spoken by a computer-generated voice. Trials could be congruent (e.g., the word "LEFT" presented to the left ear) or

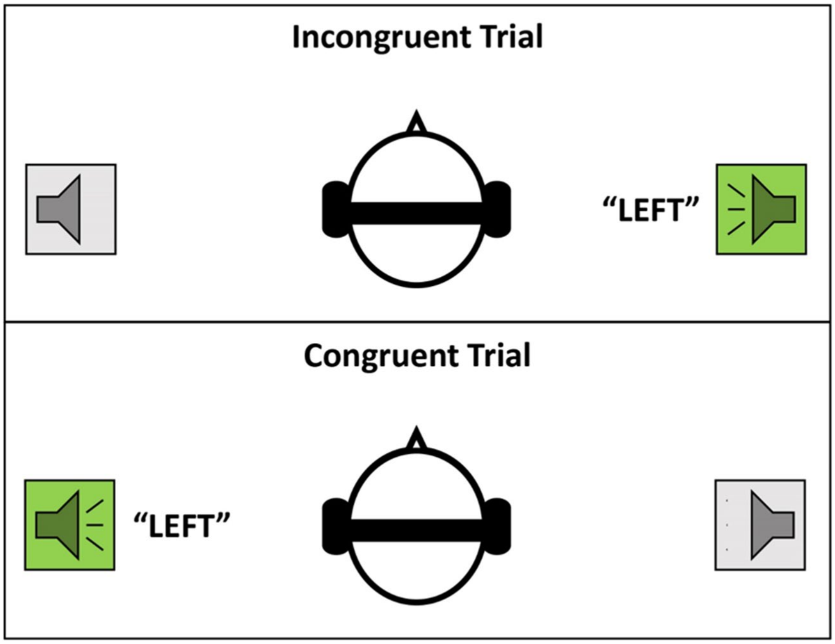

Fig. 2 Auditory Simon DL task. Top panel: example of an incongruent trial. Bottom panel: example of a congruent trial. The subject must respond based on which ear received auditory input, pressing the "P" key to indicate the right ear and the "Q" key to indicate the left ear. The challenge for the subject is to respond accurately before the adaptive response deadline while ignoring the semantic content of the auditory message

incongruent (e.g., the word "LEFT" presented to the right ear). The trial order was determined using a random seed to keep the order consistent across subjects. Trials were selected randomly on the condition that congruent trials occurred twice as often as incongruent trials by the completion of the task. Subjects completed three blocks of 96 trials, with a self-paced break between each block, for a total of 288 trials (192 of the trials were congruent and 96 were incongruent).

Subjects needed to accurately respond to the stimulus before a response deadline for the trial to be scored as correct. The response deadline was adaptive and varied according to a staircase procedure that used a 3:1 up-to-down step-size ratio based on performance on incongruent trials. Specifically, the response deadline started at 1230 ms and adapted based on whether the subject responded correctly before the response deadline: If so, the response deadline was reduced by a factor of $1x$ to make the task more difficult; if not, the subject was given the feedback "too slow" and the response deadline was increased by a factor of $3x$ to make the task easier. The value of "x" was set to 80 ms at the beginning of the task and progressively reduced to 40, 20, 10, 5, and 3 as subjects completed each sixth of the task. In other words, the step size was reduced as the task progressed. The 3:1 up-to-down adaptive procedure with two response options should converge on an accuracy rate of 75% on incongruent trials (Kaernbach, 1991). The outcome measure was the average response deadline (in milliseconds) over the final four reversals of the staircase function (i.e., trials in which the deadline either increased after it had decreased, or decreased after it had increased). Thus, the outcome measure reflects both the speed and accuracy of the participants on incongruent trials over the course of the task.

Auditory Flanker with adaptive response deadline (Auditory Flanker DL) In Auditory Flanker DL (Fig. 3), subjects must indicate whether a voice presented auditorily to the center of the headphones (i.e., presented to both ears) uttered the word "bat" or "bed" while ignoring words that were presented to just the left headphone or just the right headphone (see Chan et al., 2005, for a similar approach). Subjects were instructed to press the "P" key to indicate that the centrally presented auditory stimulus referred to a "bat" and the "Q" key to indicate that the centrally presented auditory stimulus referred to a "bed." These instructions were displayed on the screen for the duration of the task.

The auditory stimuli used in the task were created using two male computer-generated voices and two female computer-generated voices uttering the words "bat" or "bed"; each voice constituted a different version of the stimulus. On each trial, the subject was presented with one voice uttering a word to both headphones, a second voice uttering a word to the left headphone, and a third voice uttering a word to the right headphone. Voice stimuli were selected randomly,

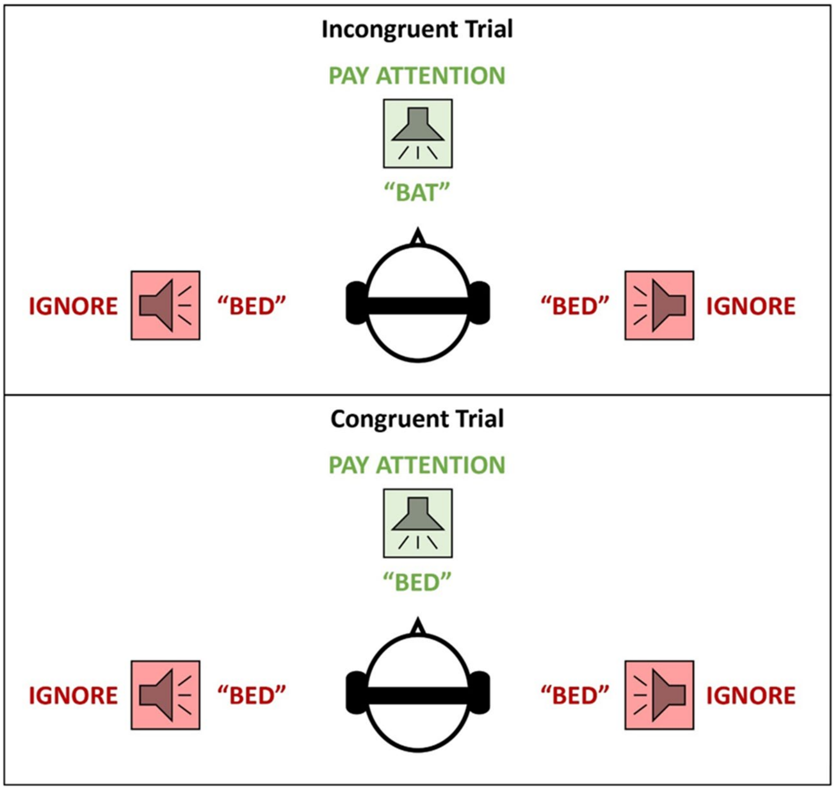

Fig. 3 Auditory Flanker DL. Top panel: example of an incongruent trial. Bottom panel: example of a congruent trial. The subject must respond based on whether the centrally presented stimulus refers to the word "bat" (i.e., press "P" key) or the word "bed" (i.e., press "Q" key). The challenge for the subject is to respond accurately before the adaptive response deadline while ignoring the flanking auditory messages

so on a given trial any three of the four voices could be presented to either or both headphones. Trials could be congruent (e.g., the word "bat" presented centrally as well as to the left ear and the right ear) or incongruent (e.g., the word "bat" presented centrally while the word "bed" was presented to the left ear and the right ear). The trial order was determined using a random seed to keep the order consistent across subjects. Trials were selected randomly on the condition that congruent trials occurred twice as often as incongruent trials by the completion of the task. Subjects completed three blocks of 96 trials, with a self-paced break between each block, for a total of 288 trials (192 of the trials were congruent and 96 were incongruent). Subjects needed to accurately respond to the stimulus before a response deadline for the trial to be scored as correct. The adaptive response deadline was programmed using the exact same parameters as those for the Auditory Simon DL task described above. The outcome measure was the average response deadline (in milliseconds) over the final four reversals of the staircase function. Thus, the outcome measure reflects both the speed and accuracy of the participants on incongruent trials over the course of the task.

Auditory Stroop with adaptive response deadline (Auditory Stroop DL) In Auditory Stroop DL (Fig. 4), subjects must indicate whether words presented auditorily to both ears referred to males or females while ignoring the tone of the voice used as the auditory stimulus. Subjects wore headphones for the test. They were shown a list of words and told which words referred to males (i.e., "brother," "dad," "father," and "boy") and which words referred to females (i.e., "sister," "mom," "mother," "girl"). They were instructed to press the "P" key to indicate that the auditory stimulus referred to a male and the "Q" key to indicate that the auditory stimulus referred to a female. These instructions were displayed on the screen for the duration of the task. The auditory stimuli consisted of the preceding list of words, spoken by either a male or female computer-generated voice. Trials could be congruent (e.g., the word "brother" presented using the male voice) or incongruent (e.g., the word "brother" presented using the female voice) (see Christensen et al., 2011, for a similar approach). The trial order was determined using a random seed to keep the order consistent across subjects. Trials were selected randomly on the condition that congruent trials occurred twice

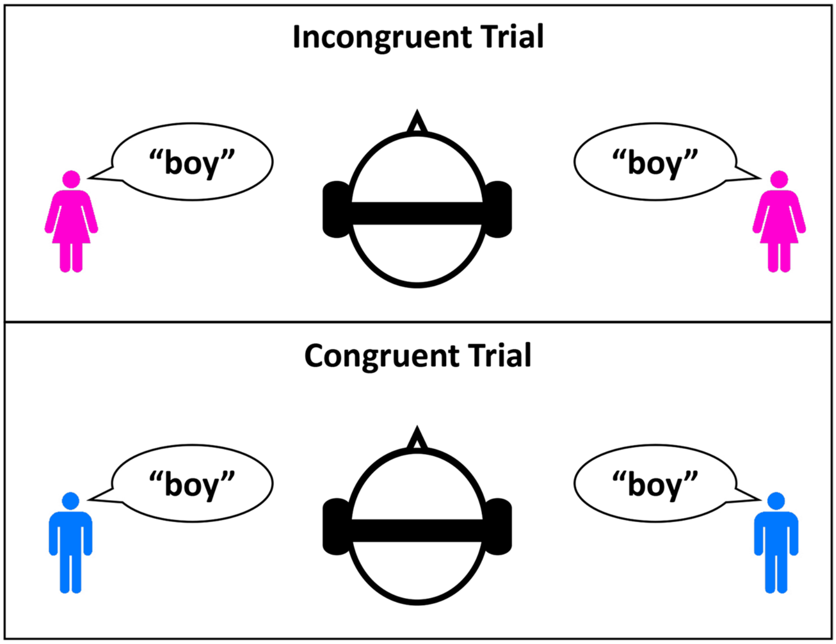

Fig. 4 Auditory Stroop DL. Top panel: example of an incongruent trial. Bottom panel: example of a congruent trial. The subject must respond based on whether the auditory stimulus refers to a male (i.e., press "P" key) or a female (i.e., press "Q" key). The challenge for the subject is to respond accurately before the adaptive response deadline while ignoring the tone of the voice used as the auditory stimulus

as often as incongruent trials by the completion of the task. Subjects completed three blocks of 96 trials, with a self-paced break between each block, for a total of 288 trials (192 of the trials were congruent and 96 were incongruent). Subjects needed to accurately respond to the stimulus before a response deadline for the trial to be scored as correct. The adaptive response deadline was programmed using the exact same parameters as those for the Auditory Simon DL task described above. The outcome measure was the average response deadline (in milliseconds) over the final four reversals of the staircase function. Thus, the outcome measure reflects both the speed and accuracy of the participants on incongruent trials over the course of the task.

Visual attention control

We used the antisaccade task (Hallett, 1978; Hutchison, 2007), the selective visual arrays task (Martin et al., 2021; Shipstead et al., 2014; Vogel et al., 2005), and the sustained-attention-to-cue task (Draheim et al., 2021; Burgoyne et al., 2023; Draheim et al., 2023; Tsukahara & Engle, under review) as indicators of visual attention control because these measures were found to be the best-performing tests in our lab's "toolbox" approach to improving the measurement of attention control (e.g., Draheim et al., 2021). Specifically, the measures demonstrated strong reliability, intercorrelations, factor loadings, and construct validity. All three tasks effectively operationalize attention control by requiring subjects to maintain focus on task-relevant information while ignoring or suppressing the influence of distractions and interference. For example, in the antisaccade task, subjects must inhibit the prepotent response of looking towards a flickering asterisk, and instead look in the opposite direction to detect a briefly presented letter. In the selective visual arrays task, subjects are shown a memory array and told to selectively attend to and remember a subset of items (e.g., remember the blue items) while ignoring the remaining items (e.g., ignore the red items). In the sustained-attention-to-cue task, subjects must remain focused on a cued spatial location on the computer screen for a variable wait period (2–12 seconds) in order to detect a briefly presented letter. These attention control measures have been tested extensively for individual differences research (for details on their psychometric properties and evidence for their construct validity, see Burgoyne et al., 2023; Draheim et al., 2021, 2023; Kane et al., 2001; Martin et al., 2021; Redick et al., 2007; Tsukahara & Engle, under review; Unsworth et al., 2004).

Antisaccade (Hallett, 1978; Hutchison, 2007) Participants identified a "Q" or "O" that appeared briefly on the opposite side of the screen as a distractor stimulus. After a central fixation cross appeared for 1000 ms or 2000 ms, an asterisk (*) flashed at 12.3° visual angle to the left or right of the central fixation for 100 ms. Afterward, the letter "Q" or "O" was presented on the opposite side at 12.3° visual angle of the central fixation for 100 ms, immediately followed by a visual mask (##). Participants indicated whether the letter was a "Q" or an "O." They completed 16 slow practice trials during which letter duration was set to 750 ms, followed by 72 test trials. The task was scored based on accuracy as the proportion of correct responses.

Sustained attention to cue (SACT; Draheim et al., 2021) The critical element in this task is the wait time interval in which attention must be sustained at a spatially cued location for a variable amount of time. After the variable wait time, a target letter is briefly presented and must be identified amidst a mix of other non-target letters. Each trial started with a central black fixation for 1 second followed by a 750 ms interval in which the words "Get Ready!" were displayed at the to-be-cued location along with an auditory beep. A circle cue was then displayed for approximately 500 ms, and then was removed from the display during the wait time interval. The wait time lasted either 0 seconds or 2–12 seconds in 500 ms intervals (e.g., 2, 2.5, 3, 3.5… seconds). After the variable wait time, a cloud array of letters was displayed at the cued location for 250 ms. The target letter was identifiable as the central letter in slightly darker font color. The target and non-target stimuli were B's, P's, or R's. The task had three blocks of 22 trials for a total of 66 trials without feedback. The task was scored as the proportion of correct responses.

Selective visual arrays (adapted from Vogel et al., 2005) Participants were shown a fixation cross for 1000 ms, followed by the word "RED" or "BLUE" that instructed them to pay attention to either the red or blue rectangles that would appear shortly. An array of red and blue rectangles arranged at different angle orientations (i.e., the "target array") appeared for 250 ms, which was followed by a blank screen lasting 900 ms. The display included three or five rectangles of each color. Afterward, an array appeared that included only the cued-color of rectangles (i.e., the "probe array"), and a white dot was used to highlight one of the rectangles. The angle of this particular rectangle could be the same as it appeared in the target array, or different; both possibilities were equally likely. The participant's task was to determine whether the angle of the rectangle was the same or had changed, using the keyboard to respond. We used 48 trials for each set size, and computed capacity scores (k) for each set size using the single-probe correction (Cowan et al., 2005): set size × (hit rate + correction rejection rate – 1). The outcome measure was the mean k estimate across set sizes 3 and 5.

Fluid intelligence

Raven's Advanced Progressive Matrices (Raven & Court, 1998) In Raven's matrices, participants were shown a grid of 3×3 line drawings patterns, with the pattern in the bottom right corner missing. The participant's task was to select from eight response options the pattern that best fit the array. We gave participants 10 minutes for 18 items from Raven's Advanced Progressive Matrices; the measure of performance was the number of items they correctly responded to.

Letter sets (Ekstrom et al., 1976) In letter sets, participants were shown five sets of four letters and challenged to identify the set of letters that did not adhere to the same pattern as the others. We gave participants 10 minutes to complete 30 items; the measure of performance was the number of items they correctly responded to.

Number series (Thurstone, 1938) In number series, we presented participants with a set of numbers that followed a pattern. They were shown four possible response options that could complete the pattern, and needed to select the response option that best followed the pattern of the number series. We gave participants 5 minutes for 15 items; the measure of performance was the number of items they correctly responded to.

Working memory capacity

To operationalize working memory capacity, we administered two complex span tests, advanced symmetry span (Unsworth et al., 2005) and advanced rotation span (Kane et al., 2004). These tasks are considered "advanced" because they include more trials at larger set sizes than the traditional versions of these tasks. As such, they more effectively distinguish performance levels among high-ability individuals (Draheim et al., 2018). We elected not to include the operation span task in the task battery because it has been shown to perform poorly when used in high-ability samples (Draheim et al., 2018). Although the other cognitive constructs in the task battery have three indicator measures, having two indicator measures for working memory capacity does not pose a problem for the latent variable analyses in the present work; all models are identified because the working memory capacity factor was allowed to correlate with the other latent cognitive ability factors (Kline, 2023).

Advanced symmetry span (Unsworth et al., 2005) In symmetry span, participants must remember spatial locations while deciding whether patterns are symmetrical or not. On a given trial, the participant was shown a symmetrical or asymmetrical grid and needed to determine whether or not it was symmetrical. Next, they were shown a 4×4 grid of squares, and one of them was emphasized by a red color. Their goal was to memorize the location of the colored square. This symmetry/square interleaving pattern continued two to seven times (i.e., the set sizes used in the task).

Afterward, the participant needed to report the location that the colored squares appeared in, in the order that they appeared. We gave participants 12 trials, two of each set size. We used the partial scoring method as the outcome measure of performance.

Advanced rotation span (Kane et al., 2004) In rotation span, participants remembered directional arrows while deciding whether a letter was in the proper orientation or mirror-imaged. On a given trial, the participant was shown a letter they would mentally rotate to determine its orientation (mirror-imaged or normal). Next, they were shown a single arrow that was either small or large and pointed in one of eight directions. This letter/arrow interleaving pattern continued two to seven times (i.e., the set sizes used in the task). Afterward, the participant was asked to report the arrows in the order they appeared. We gave participants 12 trials, two of each set size. We used the partial scoring method as the outcome measure of performance.

Processing speed

Digit string comparison (Redick et al., 2012) Participants were shown three, six, or nine numbers that appeared on the left and right side of a horizontal line drawn between them. The participant's task was to determine whether the strings of digits were identical or different. They responded using the mouse. Participants were given two blocks of 30 seconds of trials and attempted to answer as many items correctly as possible. Participants earned one point for each correct response and lost one point for each incorrect response; the measure of performance was the number of points earned at the conclusion of the task.

Letter string comparison (Redick et al., 2012; Salthouse & Babcock, 1991) This task was almost identical to the digit string comparison task, however, instead of digits, the participant made comparisons about strings of letters.

Pattern comparison (Salthouse & Babcock, 1991) The participant was shown two symbols that appeared on either side of a horizontal line and indicated whether they were the same or different. Participants were given two blocks of 30 seconds of trials and attempted to answer as many items correctly as possible. Participants earned one point for each correct response and lost one point for each incorrect response; the measure of performance was the number of points earned at the conclusion of the task.

Criterion measures

Dichotic listening Participants completed two dichotic listening tests from the Performance Based Measures, part of the Aviation Selection Test Battery used by the U.S. Navy, Marine Corps, and Coast Guard for job classification (Walker et al., 2007). Participants were presented with a series of numbers and letters via headphones to each ear and instructed to monitor a target ear while ignoring the other one. Their task was to press the trigger of a joystick when they heard an even number and to press the thumb button of a throttle when they heard an odd number. Performance was measured during the isolated dichotic listening test as well as during a multitask in which participants performed dichotic listening while completing a two-handed tracking task. A composite score was computed based on both dichotic listening tests that accounted for the speed and accuracy of responses. Because the method of scoring the Performance Based Measures is proprietary, we could not extract a measure of isolated dichotic listening performance from the raw data. Thus, we use the composite score as a criterion measure in this work, noting that it represents complex cognitive performance with auditory and visual attentional demands. See Walker et al. (2007) and Mashburn et al. (under review) for additional details.

Foster multitask (Martin et al., 2020; Fig. 5) In this multitask, subjects must coordinate performance on four visual-modality subtasks. The four subtasks included mathematics, word recall, and two visual monitoring subtasks (i.e., a battery monitoring task and a spinning disc monitoring task). The outcome measure was the average score across three 5-minute test blocks. Task details are presented in Martin et al. (2020).

Transparency and openness

We report all data exclusions below. This study's design and its analysis were not pre-registered. Data, task downloads, and R scripts are openly available on the Open Science Framework (https://osf.io/2zqe7/). Data were collected as part of a larger project, the details of which are provided online (https://tinyurl.com/ms8htaep).

Data preparation

We removed participants' scores on a task if they showed severely poor performance indicating that they did not understand the instructions or were not performing the task as intended. Specifically, we computed chance-level performance on each task; any scores that were at or below chance-level performance were identified as problematic data points and set to missing. This procedure was applied to the three auditory attention control tasks, the three visual attention control tasks, and the three processing speed tasks. For the two working memory tasks, problematic data points were defined by chance-level performance or worse on the

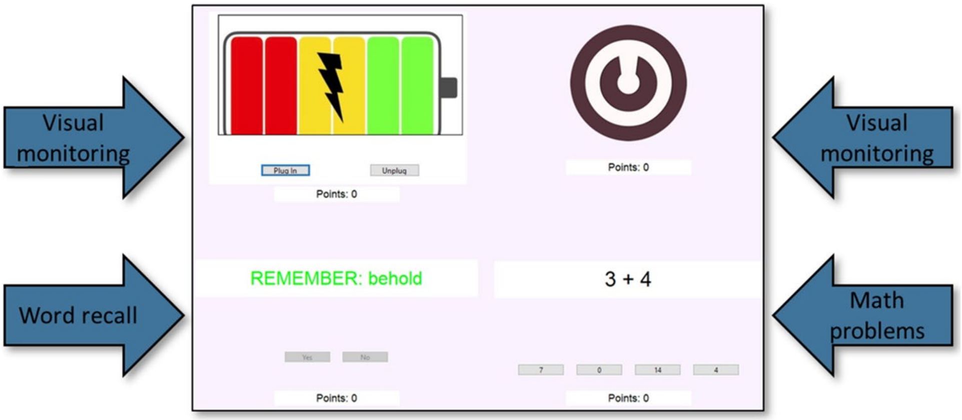

Fig. 5 A labeled snapshot of the Foster multitask interface

processing subtask. We did not remove problematic data points for the three fluid intelligence tests, dichotic listening, or multitasking paradigms, as scores of zero are possible even if subjects understand the instructions. After removing problematic data points, we performed a two-pass outlier exclusion procedure for all tasks. We removed data points that were more than 3.5 standard deviations worse than the sample mean two times, recomputing the sample mean and standard deviation each time. Altogether, this led to the exclusion of less than 5% of the data for each task.[1]

Modeling approach and fit statistics

For all confirmatory factor analyses and structural equation models, we used maximum likelihood estimation with robust standard errors and full-information maximum likelihood estimation for missing data. Variables were standardized prior to estimation. Confirmatory factor analyses and structural equation models were estimated using JASP [Jeffreys's Amazing Statistics Program] (JASP Team, 2023).

We report multiple fit statistics: The χ^2 is an absolute fit index comparing the fit of the specified model to that of the observed covariance matrix. A significant χ^2 can indicate lack of fit, but is heavily influenced by sample size. In large samples, such as the one used in the present studies, even a slight deviation between the data and the model can lead to a significant χ^2 statistic. Therefore, we also report the comparative fit index (CFI) and Tucker–Lewis index (TLI), which compare the fit of the model to a null model in which the covariation between measures is set to zero, while adding penalties for additional parameters. For CFI and TLI, large values indicate better fit (i.e., > .90 or ideally, > .95). For the root-mean-square error of approximation (RMSEA) fit statistic, values less than .05 are considered excellent, while values less than .10 are considered only adequate. For the standardized root-mean-square residual (SRMR), which computes the standardized difference between the observed and predicted correlations, a value of less than .08 indicates adequate fit (Hu & Bentler, 1999).

Results

Descriptive statistics

Demographic information is provided in Table 1. The participants' average age was 22 (*SD* = 4) years, and a majority were female (59%). In terms of race/ethnicity, 41% of the sample identified as Asian or Pacific Islander, 28% identified as White, 13% identified as Black or African American, and the remainder selected "other" or declined to respond. The majority of participants (91%) had attended at least some college.

Descriptive statistics are reported in Table 2. On average, the auditory attention control tasks required between 5 and 8 minutes of testing time, whereas the visual attention control tasks required between 12 and 17 minutes of testing time.

We explored multiple approaches to computing internal consistency on the auditory attention control tests. Because of the adaptive response deadline, measures of performance across trials are not independent of one another. Thus, all of these estimates of reliability should be seen as

---

[1] Specific details regarding removed cases for the Auditory DL tasks are as follows: For Auditory Flanker DL, there was an initial sample of 312 subjects; 3 were identified as problematic, 3 were first-pass outliers, and 5 were second-pass outliers. For Auditory Simon DL, there was an initial sample of 314 subjects; 3 were identified as problematic, 5 were first-pass outliers, and 6 were second-pass outliers. For Auditory Stroop DL, there was an initial sample of 317 subjects; 0 were identified as problematic, 6 were first-pass outliers, and 1 was a second-pass outlier.

Table 1 Demographic information

Demographic	Statistic
Age (years)	Mean = 21.95 *SD* = 4.09 Range = 18 – 35
Gender	Male = 39.5% Female = 58.9% Self-identify/other = 1.3% Transgender male = 0.3%
At least some college?	Yes = 90.8% No = 9.2%
Race/Ethnicity	White = 28.3% Black or African American = 13.4% Asian or Pacific Islander = 41.4% Other* = 16.9%

*Other includes Hispanic or Latino, Native American, and others. *N* = 314. Due to experimenter error, a small number of participants did not complete the demographic questionnaire

Table 2 Descriptive statistics

Measure	*N*	*M*	*SD*	Skew	Kurtosis	Reliability	Time (*SD*)
Auditory Simon DL	300	1003.00	516.45	1.74	2.76	.92$^{\alpha}$	5.29 (1.19)
Auditory Flanker DL	301	1306.55	648.58	1.60	2.58	.90$^{\alpha}$	7.74 (1.60)
Auditory Stroop DL	310	1104.76	301.58	1.26	1.16	.89$^{\alpha}$	6.01 (0.96)
Antisaccade	299	0.81	0.12	−0.62	−0.64	.87$^{\alpha}$	---
SACT	307	0.89	0.10	−1.11	0.71	.87$^{\alpha}$	17.11 (1.50)
Selective visual arrays	316	2.47	0.70	−0.51	0.07	.81^{b}	12.05 (2.01)
Raven's matrices	316	11.30	2.87	−0.41	−0.26	.77$^{\alpha}$	---
Letter sets	312	16.41	4.41	−0.17	−0.69	.85$^{\alpha}$	---
Number series	317	9.99	2.98	−0.22	−0.73	.73$^{\alpha}$	---
Symmetry span	310	29.90	9.73	−0.24	−0.40	.76$^{\alpha}$	---
Rotation span	310	25.16	8.66	−0.11	−0.20	.73$^{\alpha}$	---
Digit comparison	307	29.90	5.51	−0.45	0.02	.88^{b}	---
Letter comparison	307	20.53	4.10	0.12	0.39	.82^{b}	---
Pattern comparison	306	39.06	6.01	−0.09	−0.22	.94^{b}	---
Dichotic listening	260	24.82	10.54	−0.53	−0.62	---	---
Foster multitask	302	96011.66	26343.41	−0.20	0.05	.95$^{\alpha}$	---

$^{\alpha}$ Cronbach's alpha, ^{b} split-half reliability with Spearman–Brown correction. Time (*SD*) = mean administration time in minutes (standard deviation); --- = administration time or reliability was not measured or could not be computed. For a comparison of performance on incongruent and congruent trials in the auditory conflict tasks, see Table 5

approximations, and are likely inflated due to within-task dependencies. First, we computed Cronbach's alpha across the final four reversals of the staircase procedure, because these final four reversal points were averaged when computing participants' final scores on the task. Cronbach's alpha revealed excellent internal consistency for Auditory Simon DL (α = .92), Auditory Flanker DL (α = .90), and Auditory Stroop DL (α = .89). We also computed the correlation between the response deadline at the midpoint of the task and at the end of the task. These correlations revealed good internal consistency reliability as well: Auditory Simon DL ($r = .99$, $p < .001$), Auditory Flanker DL ($r = .99$, $p < .001$), and Auditory Stroop DL ($r = .97$, $p < .001$). Finally, we computed correlations across different methods of scoring the auditory conflict tasks. The measure of performance used in the present work is the response deadline over the final four reversals of the staircase procedure; we correlated this measure with deadlines formed by averaging over the last eight reversals, the last 10 reversals, and over all reversals, and by taking the response deadline at the conclusion of the task. For all three auditory attention control tasks, these correlations exceeded $r = .98$. Thus, the measure of performance seems fairly robust to alternative scoring methods involving the response deadline. Nevertheless, we reiterate that dependencies across trials in the tasks likely inflate these estimates.

We computed Cronbach's alpha on the three visual attention control tasks: α = .87 for antisaccade, α = .87 for the SACT, and α = .81 for selective visual arrays. Note that because trials were independent in the visual tasks but not independent in the auditory tasks, readers should be cautious when comparing Cronbach's alpha for the two sets of attention control tasks.

The three auditory attention control tests had estimates of skewness ranging from 1.26 to 1.74, and estimates of kurtosis ranging from 1.16 to 2.76. Inspection of the histograms revealed right-skewed distributions for all three

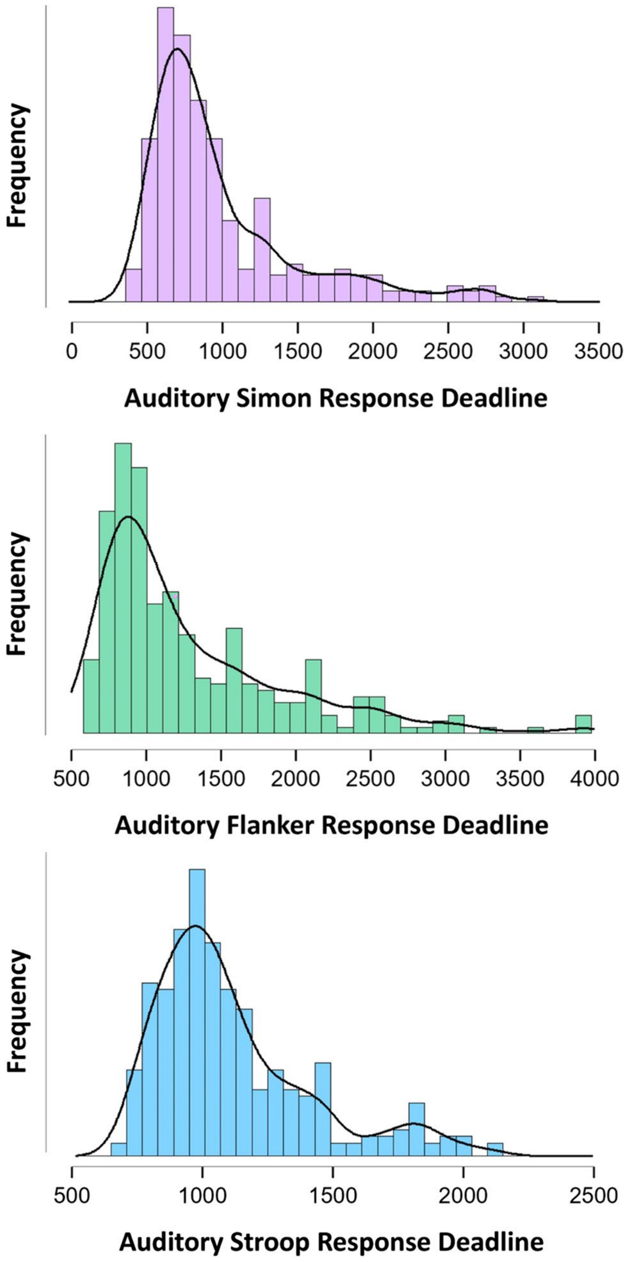

Fig. 6 Histogram of scores on Auditory Simon DL (top panel; *n* = 300), Auditory Flanker DL (middle panel; *n* = 301), and Auditory Stroop DL (lower panel; *n* = 310)

auditory tasks (see Fig. 6). We therefore performed a log transformation on the three auditory attention control measures, which resulted in estimates of skewness and kurtosis between −1 and +1. We used the log-transformed auditory attention control measures in the correlational and factor analytic analyses reported below.

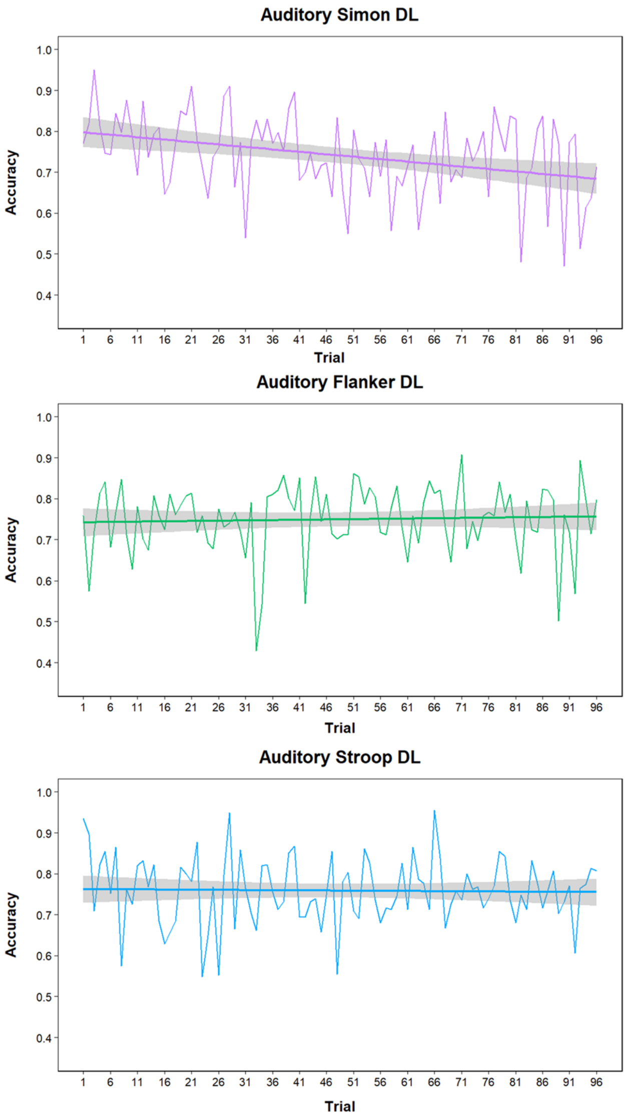

Fig. 7 Converged upon accuracy rate on incongruent trials on Auditory Simon DL (top panel), Auditory Flanker DL (middle panel), and Auditory Stroop DL (lower panel). Error bars represent 95% confidence intervals. Observations were nested within participants. The *x*-axis represents the incongruent trial number

Adaptive response deadline convergence

Across the sample, the auditory tasks yielded an average response deadline of 1.0 seconds for Auditory Simon DL, 1.1 seconds for Auditory Stroop DL, and 1.3 seconds for Auditory Flanker DL (Table 2). However, these sample averages belie a large amount of variability in the response deadline across subjects, as shown in the histograms in Fig. 6. Final response deadlines ranged from as little as 500 ms to nearly 4 seconds across subjects and tasks. Given this variability, we tested whether the staircase procedure was successful in identifying the response deadline at which the subject could respond correctly to incongruent trials 75% of the time, for a given task (Kaernbach, 1991).

Figure 7 shows the converged-upon accuracy rate on incongruent trials for the three auditory tasks, based on mixed-effects models with observations nested within participants. For Auditory Flanker DL and Auditory Stroop DL, the figures clearly show an accuracy rate of approximately 75% that varied only slightly over the course of the task. For Auditory Simon DL, the accuracy rate started around 80% and drifted towards 70% as the task progressed. We computed the mean accuracy rate on incongruent trials for each of the three tasks and found that it was 74.10% ($SD = 5.88\%$) for Auditory Simon DL, 74.92% ($SD = 6.53\%$) for Auditory Flanker DL, and 75.86% ($SD = 4.12\%$) for Auditory Stroop DL. Thus, the adaptive staircase procedure effectively obtained an accuracy rate of approximately 75% on incongruent trials when averaging across participants.

Next, we tested whether differences in the ability to control attention on the visual attention control tasks (i.e., antisaccade, selective visual arrays, and SACT) predicted performance on the auditory attention control tasks. We created a composite measure by averaging z scores (i.e., standardized scores) on the three visual attention control tasks. We then estimated the response deadline for individuals at $z = -2$, $z = 0$, and $z = 2$, to illustrate how subjects at -2 *SD*, the mean, and 2 *SD* on the visual attention control tasks performed on the auditory attention control tasks. As shown in Fig. 8, there was an effect of visual attention control on performance in the auditory tasks, with subjects who ranked higher on visual attention tasks demonstrating quicker response deadlines on Auditory Simon DL, Auditory Flanker DL, and Auditory Stroop DL. These analyses provide the first evidence of construct validity for the auditory attention control measures, which we examined in depth using correlational and factor-analytic methods.

Convergent and divergent validity of the auditory conflict tasks

As shown in Table 3, the auditory attention control measures had significant intercorrelations ranging from $r = .22$ to $r = .33$, with an average of $r = .28$. Furthermore, the auditory attention control measures correlated significantly with the visual attention control measures (average $r = .25$). This indicates that the auditory and visual attention control measures tap overlapping sources of variance.

We investigated the factor structure of the cognitive ability measures by first conducting an exploratory factor analysis. We entered the cognitive ability measures into the analysis, withholding the criterion measures (i.e., dichotic listening and multitasking) because they reflect performance on complex cognitive tests that likely invoke more than one broad cognitive ability. We extracted four factors with eigenvalues greater than 1 using principal axis factoring with an oblique oblimin rotation; factor loadings are reported in Table 4. The six attention control measures clustered on Factor 1, the three fluid intelligence measures clustered on Factor 2, the three processing speed measures clustered on Factor 3, and the two complex span measures of working memory capacity clustered on Factor 4. There were a few moderate-in-size cross-loadings, such as selective visual arrays loading on both the attention control factor (.29) and the working memory capacity factor (.38), but in general, the results are broadly consistent with our a priori expectations about the factor structure of the cognitive ability measures.[2] The observation that the auditory and visual attention control measures loaded highly on a common factor provides supporting evidence for the construct validity of the auditory tasks and the domain-generality of the attentional ability that they measure.

Next, we conducted a series of confirmatory factor analyses. In our first model, we specified a single attention control factor with six indicators, including the three measures of auditory attention control as well as the three measures of visual attention control. The model is depicted in Fig. 9 and fit the data well. The auditory attention control measures had loadings ranging from .43 to .49, and the visual attention control measures had loadings ranging from .51 to .66. Thus, consistent with the results of the exploratory factor analysis reported above, the auditory and visual attention control measures loaded well on a common attention control factor, providing support for the modality-general view of attention control.

In our second model, we specified separate latent factors for auditory attention control and visual attention control and estimated the proportion of variance shared between them (see Fig. 10). Factor loadings for the auditory tasks ranged from .52 to .56; loadings for the visual tasks ranged from .52 to .68. Crucially, the correlation between the two factors was $r = .81$, indicating that the two factors shared approximately 66% of their reliable variance (i.e., $.81^2$). Setting this correlation equal to 1, however, significantly reduced model fit ($\Delta\chi^2\ (1) = 6.62$, $p = .010$). This indicates that the auditory and visual attention control measures share a majority of their reliable variance but that the two constructs are not perfectly isomorphic with one another.

In our third model, we investigated the latent relationships between the auditory attention control measures, visual attention control measures, fluid intelligence, working memory capacity, and processing speed. We specified a five-factor confirmatory factor analysis model which is depicted

---

[2] For an in-depth analysis of selective visual arrays as a measure of attention control and working memory capacity, please see Martin et al. (2021).

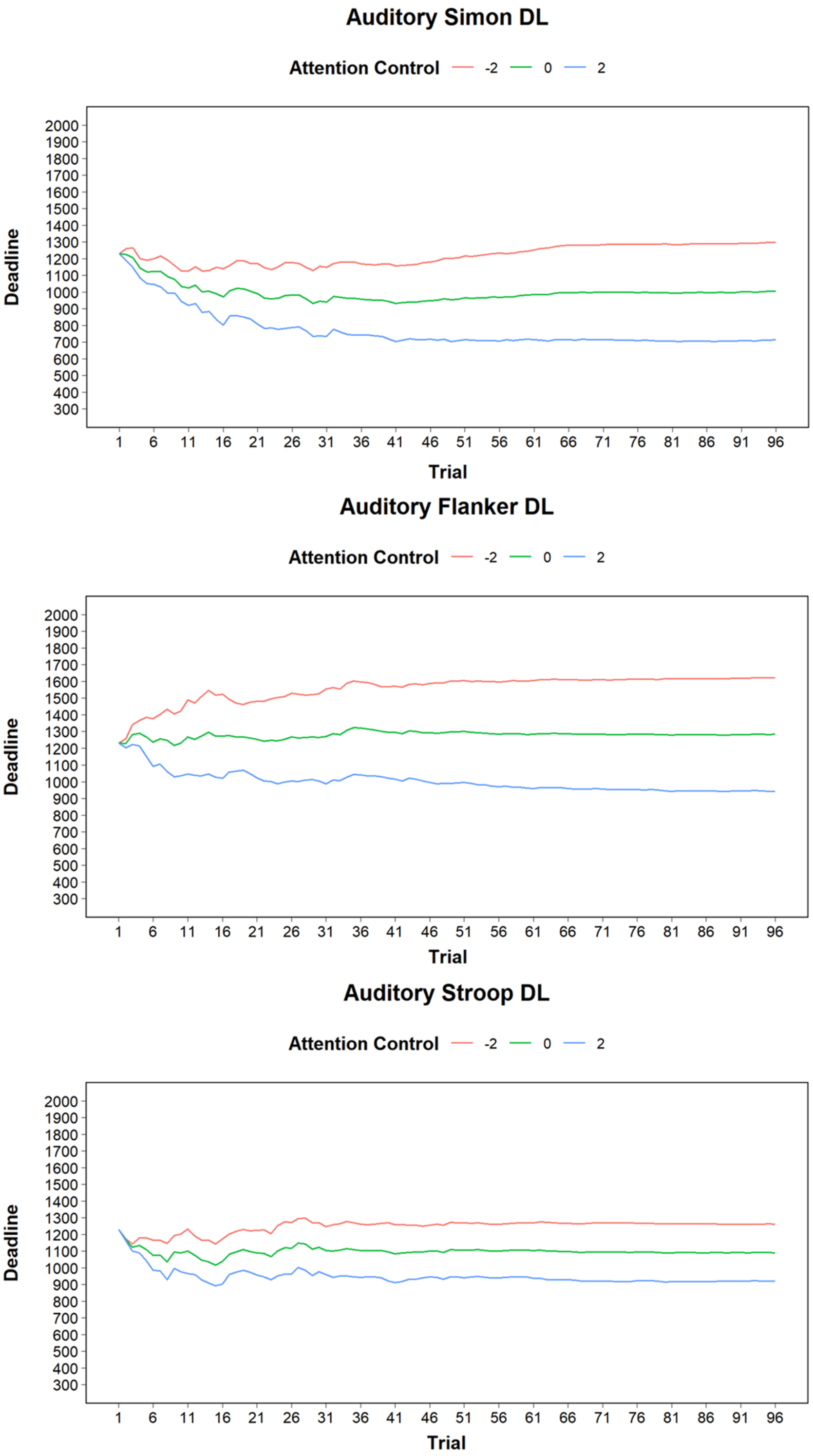

Fig. 8 Predicted response deadlines for Auditory Simon DL (top panel), Auditory Flanker DL (middle panel), and Auditory Stroop DL (lower panel) based on *z*-score average of visual attention control tasks (i.e., antisaccade, selective visual arrays, and SACT). The *x*-axis represents the incongruent trial number

in Fig. 11. Importantly, the auditory attention control factor correlated significantly more strongly with the visual attention control factor ($r = .76$, $p < .001$) than it did with fluid intelligence ($r = .48$, $p < .001$; difference = .28, $Z = 3.45$, $p < .001$), working memory capacity ($r = .35$, $p < .001$; difference = .41, $Z = 4.36$, $p < .001$), and processing speed ($r = .45$,

Table 3 Correlations between measures of cognitive ability

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Auditory Simon DL	--														
2. Auditory Flanker DL	**.30**	--													
3. Auditory Stroop DL	**.22**	**.33**	--												
4. Antisaccade	**.30**	**.28**	**.30**	--											
5. SACT	**.29**	**.18**	**.21**	**.30**	--										
6. Selective visual arrays	**.20**	**.26**	**.23**	**.39**	**.32**	--									
7. Raven’s matrices	**.15**	**.20**	.09	**.29**	.11	**.44**	--								
8. Letter sets	**.18**	**.23**	**.18**	**.26**	.09	**.32**	**.40**	--							
9. Number series	**.17**	**.23**	**.18**	**.27**	.06	**.44**	**.45**	**.58**	--						
10. Symmetry span	.10	**.17**	.07	**.23**	**.14**	**.43**	**.33**	**.30**	**.31**	--					
11. Rotation span	**.14**	**.15**	**.22**	**.27**	**.14**	**.33**	**.29**	**.19**	**.29**	**.51**	--				
12. Digit comparison	**.14**	**.21**	**.20**	**.31**	**.19**	**.34**	**.28**	**.44**	**.40**	**.24**	**.24**	--			
13. Letter comparison	.07	.09	.11	**.21**	**.16**	**.26**	**.19**	**.43**	**.30**	**.23**	**.19**	**.61**	--		
14. Pattern comparison	**.21**	**.22**	**.16**	**.30**	**.21**	**.43**	**.39**	**.32**	**.37**	**.33**	**.32**	**.49**	**.41**	--	
15. Dichotic listening	**.31**	**.28**	**.30**	**.31**	**.28**	**.31**	**.22**	**.21**	**.26**	**.22**	**.19**	**.30**	**.20**	**.29**	--
16. Foster multitask	**.27**	**.33**	**.29**	**.38**	**.26**	**.48**	**.39**	**.52**	**.63**	**.34**	**.28**	**.58**	**.47**	**.50**	**.40**

Bold, $p < .05$. Pairwise $n = 241–312$. DL = adaptive response deadline. Some correlations with the auditory attention control measures were multiplied by −1 so that positive correlations reflect better performance on each pair of measures

Table 4 Exploratory factor analysis

Measure	Factor 1 (AC)	Factor 2 (Gf)	Factor 3 (PS)	Factor 4 (WMC)	Uniqueness
Auditory Simon DL	**−.54**	−.04	.05	.06	.73
Auditory Flanker DL	**−.50**	−.17	.08	.05	.71
Auditory Stroop DL	**−.50**	−.04	.00	.06	.75
Antisaccade	**.50**	.05	.07	.13	.63
SACT	**.50**	−.24	.14	.09	.71
Selective visual arrays	**.29**	.17	.04	**.38**	.53
Raven’s matrices	.06	**.44**	−.06	.28	.61
Letter sets	.03	**.65**	.20	−.07	.45
Number series	.01	**.76**	−.01	.07	.37
Symmetry span	−.07	.04	.02	**.74**	.45
Rotation span	.09	−.03	.02	**.60**	.60
Digit comparison	.11	.11	**.68**	.00	.38
Letter comparison	−.06	−.02	**.83**	.01	.34
Pattern comparison	.16	.09	**.34**	.27	.59
Eigenvalue	1.61	1.56	1.52	1.47	

Principal axis factoring with oblimin rotation. Loadings in bold are greater than .28. Gf = fluid intelligence; AC = attention control; PS = processing speed; WMC = working memory capacity. Correlations between factors: AC with Gf = .38; AC with PS = .33; AC with WMC = .39; Gf with PS = .49; Gf with WMC = .47; PS with WMC = .35. Model fit: χ^2 (41) = 68.50, $p = .005$; CFI = .976, TLI = .946, RMSEA = .046, 90% CI [.026, .064], SRMR = .026, BIC = −168.13. Factor loadings for the auditory tasks on the attention control factor are negative because the outcome measure is defined such that lower values (i.e., response deadlines) represent better performance. A scree plot is provided in the Appendix (Fig.19)

$p < .001$; difference = .31, $Z = 3.77$, $p < .001$). This provides further evidence for the modality-generality of attention control because it shows that auditory attention control was more closely related to visual attention control than any of the other cognitive constructs represented in the model.

The role of processing speed

It could be argued that the auditory and visual attention control factors correlate with one another because they mutually depend on the speed with which individuals can process

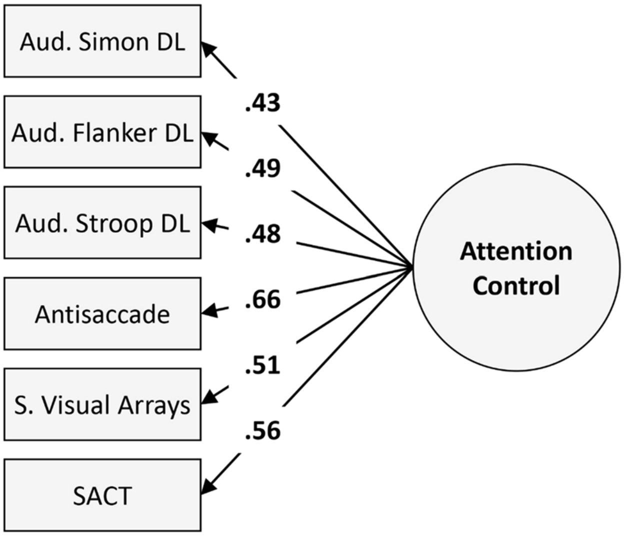

Fig. 9 Confirmatory factor analysis with three auditory and three visual attention control measures specified as indicators of a common attention control latent. Model fit: χ^2 (9) = 17.67, p = .039; CFI = .962, TLI = .937, RMSEA = .055, 90% CI [.012, .092], SRMR = .034. Loadings for the auditory attention control indicators were multiplied by −1 so that positive loadings reflect better performance

perceptual information and come to a decision and, for the auditory tasks, quickly respond via the keyboard. We tested this using perceptual processing speed tasks that require speeded encoding, decision-making, and motor response execution. We reasoned that if these sources of variance give rise to the positive correlation between auditory and visual attention control factors, then controlling for them should reduce or eliminate the correlation between auditory and visual attention control. Using latent variable modeling, we specified a processing speed factor as a predictor of auditory and visual attention control, and let the residuals of the auditory and visual attention control factors correlate with one another. These residuals (or disturbance terms) capture the variance in the attention control factors that is not attributable to perceptual processing speed. As shown in Fig. 12, after partialing out variance attributable to processing speed, the correlation between visual and auditory attention control factors was only reduced from $r = .76$ to $r = .75$ (i.e., compare the correlations in Figs. 11 and 12). Despite the fact that processing speed was a significant predictor of visual attention control (β = .63, $p < .001$; $R^2 = 39.69\%$) and auditory attention control (β = .43, $p < .001$; $R^2 = 18.49\%$), controlling for this shared variance did not diminish the relation between the visual and auditory attention control factors. In other words, the vast majority of the variance that the visual and auditory attention control factors share is not due to differences in the speed with which subjects could make simple perceptual judgments and responses.

Criterion validity: Modality matters

In our final set of analyses, we estimated the validity of auditory attention control and visual attention control for predicting two theoretically relevant outcomes: dichotic listening and multitasking performance. We first examined dichotic listening performance. As we noted in the Method section, the dichotic listening measure comprises an isolated dichotic listening test as well as a multitask in which dichotic listening is paired with two-handed tracking of moving targets. Thus, the measure represents complex cognitive performance with both auditory and visual attentional demands, although only responses to the auditory stimuli factored into participants' dichotic listening scores.

Using confirmatory factor analysis, we established that dichotic listening performance was significantly correlated with both auditory attention control ($r = .57$, $p < .001$) and visual attention control ($r = .50$, $p < .001$) at the latent level. These two correlations were not statistically significantly different from one another (difference = .07, $Z = 0.86$, $p = .388$); both visual and auditory attention control were strongly correlated with dichotic listening performance (Fig. 13).

Nevertheless, in a structural equation model with auditory attention control and visual attention control specified as correlated predictors of dichotic listening performance (Fig. 14), only the unique contribution of auditory attention control was statistically significant. That is, the unique contribution of visual attention control above and beyond auditory attention control was not significant (β = .12, $p = .605$), whereas the unique

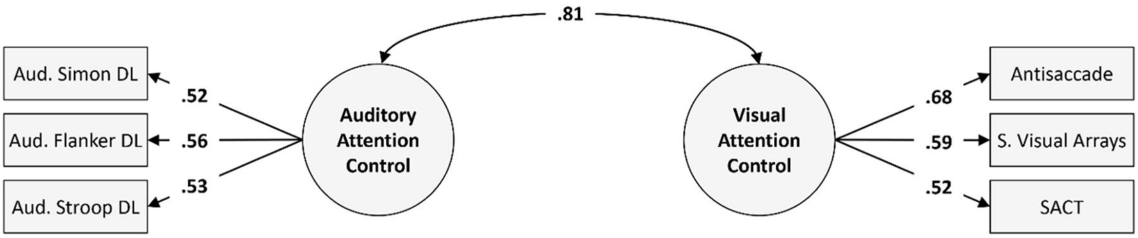

Fig. 10 Confirmatory factor analysis with latent factors representing auditory attention control and visual attention control. Model fit: χ^2 (8) = 11.05, p = .199; CFI = .987, TLI = .975, RMSEA = .034, 90% CI [.000, .079], SRMR = .027. The correlation between latent factors was multiplied by −1 so that a positive value reflects better performance on both latent factors

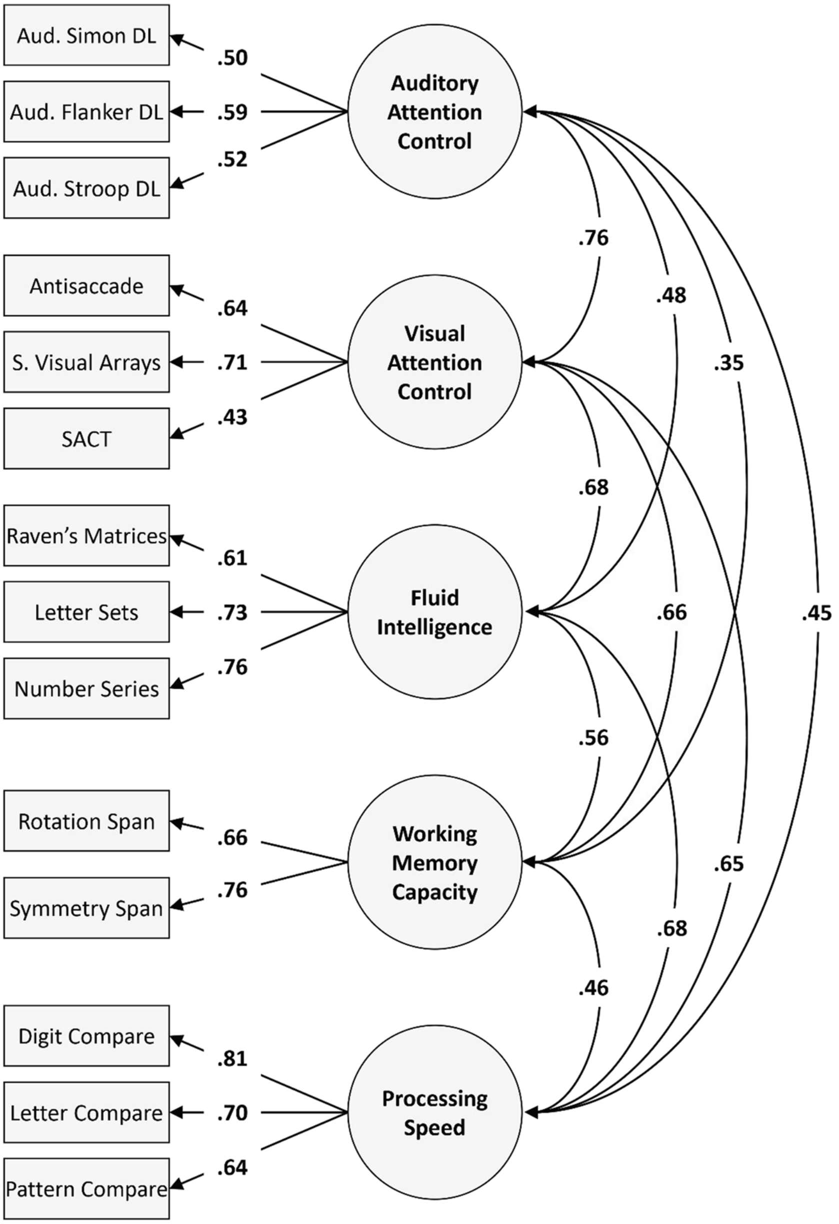

Fig. 11 Confirmatory factor analysis with latent factors representing auditory attention control, visual attention control, fluid intelligence, working memory capacity, and processing speed. Model fit: χ^2 (67) = 145.32, $p < .001$; CFI = .929, TLI = .904, RMSEA = .060, 90% CI [.047, .074], SRMR = .052. Correlations with the auditory attention control factor were multiplied by −1 so that positive correlations reflect better performance

contribution of auditory attention control above and beyond visual attention control was significant (β = .47, $p = .038$). Together, the two attention control factors accounted for 32.8% of the variance in dichotic listening performance. In subsequent models, we found that auditory attention control accounted for 31.6% of the variance on its own, whereas visual attention control accounted for 24.3% of the variance on its own.

We performed a commonality analysis to determine the unique and shared contributions of each attention factor (Nimon et al., 2010). Auditory attention control uniquely explained 8.5% of the variance in dichotic listening, visual attention control uniquely explained 1.2% of the variance, and auditory and visual attention control shared 23.1% of the variance in dichotic listening performance (see Fig. 15). Our interpretation of this result is that the modality matters: although both sets of attention control tasks were related to dichotic listening performance, auditory attention control captured variance unique to dichotic listening that visual attention control did not. We also note that the predictors accounted for a substantial amount of overlapping variance in dichotic listening

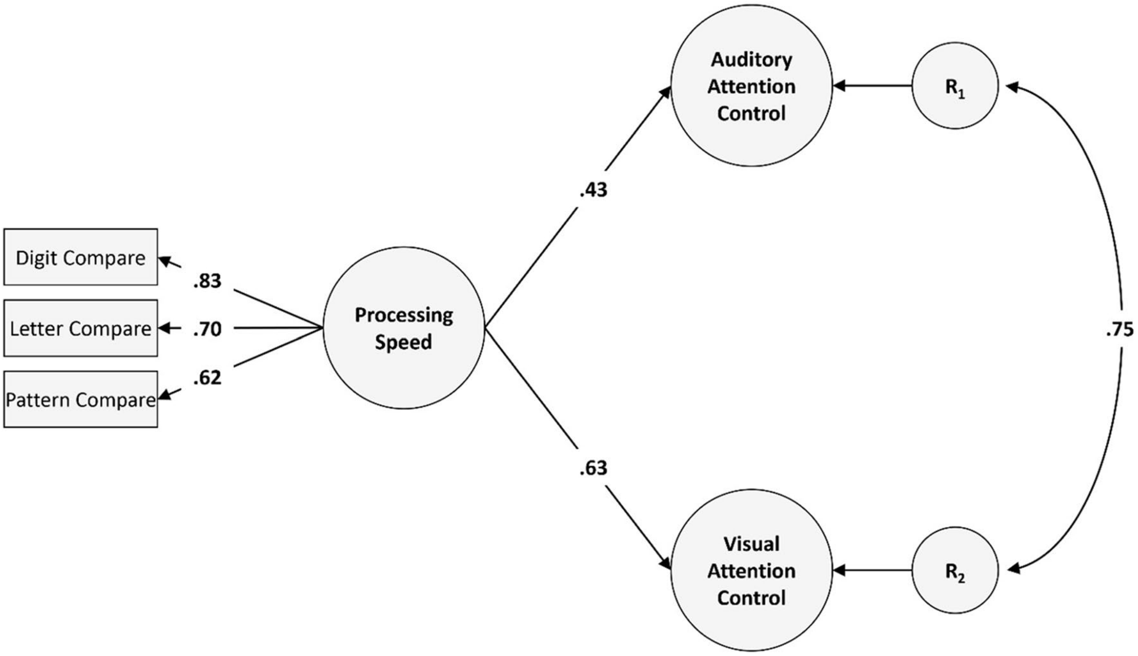

Fig. 12 Structural equation model with processing speed predicting the auditory and visual attention control factors. The residuals of the auditory and visual attention control factors remained significantly correlated ($r = .75$, $p < .001$). Indicator measures for the auditory and visual attention control factors are not shown for visual clarity. χ^2 (24) = 44.46, $p = .007$; CFI = .962, TLI = .943, RMSEA = .052, 90% CI [.027, .075], SRMR = .045

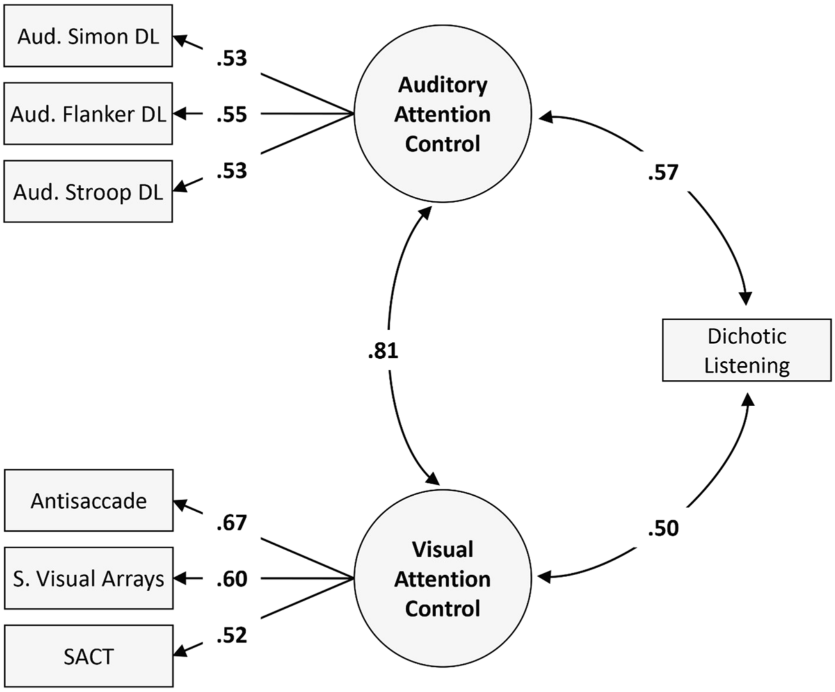

Fig. 13 Confirmatory factor analysis depicting correlations between an auditory attention control factor, a visual attention control factor, and dichotic listening performance. χ^2 (12) = 11.82, $p = .460$; CFI = 1.00, TLI = 1.00, RMSEA = .000, 90% CI [.000, .056], SRMR = .024

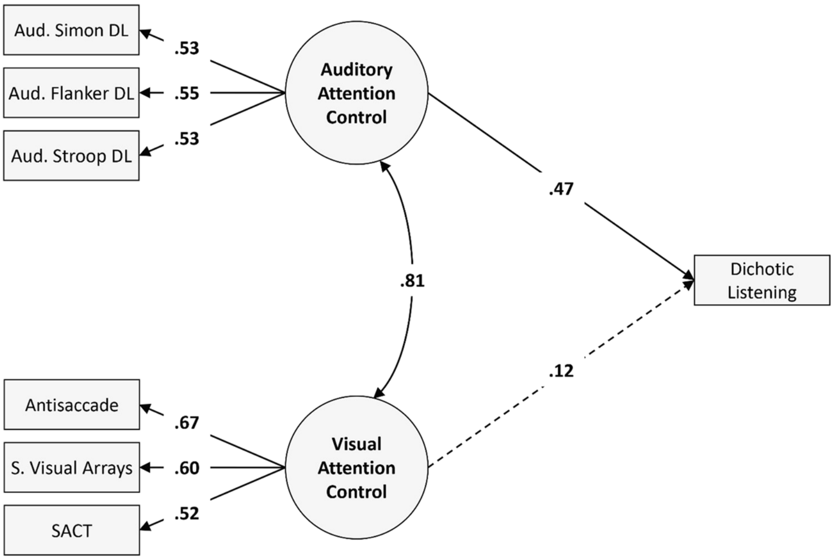

Fig. 14 Structural equation model with an auditory attention control factor and a visual attention control factor predicting dichotic listening performance. χ^2 (12) = 11.82, p = .460; CFI = 1.00, TLI = 1.00, RMSEA = .000, 90% CI [.000, .056], SRMR = .024

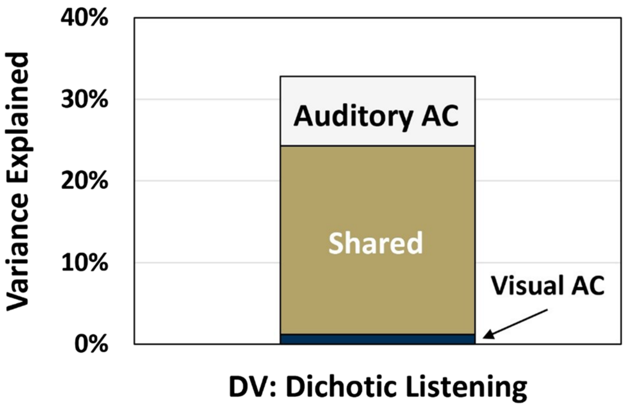

Fig. 15 Commonality analysis decomposing variance in dichotic listening performance

performance, suggesting an important role of modality-general attention control. Therefore, we see evidence for modality-generality and modality-specificity from these results.

Next, we investigated individual differences in multitasking performance. We used the Foster multitask as an indicator of multitasking performance because it contains four subtasks that are based on the visual modality. Confirmatory factor analysis revealed that multitasking performance was significantly correlated with both auditory attention control (r = .56, p < .001) and visual attention control (r = .66, p < .001). These two correlations were not statistically significantly different from one another (difference = .10, Z = 1.35, p = .176; see Fig. 16).

In a structural equation model with auditory attention control and visual attention control specified as correlated predictors of multitasking performance (Fig. 17), only the unique contribution of visual attention control was statistically significant. That is, the unique contribution of auditory attention control above and beyond visual attention control was not significant (β = .12, p = .486), whereas the unique contribution of visual attention control above and beyond auditory attention control was significant (β = .57, p < .001). Together, the two attention control factors accounted for 43.9% of the variance in multitasking performance. In subsequent models, we found that auditory attention control accounted for 30.7% of the variance on its own, whereas visual attention control accounted for 43.5% of the variance on its own.

Once again, we performed a commonality analysis to determine the unique and shared contributions of each attention factor (Nimon et al., 2010). Auditory attention control uniquely explained 0.4% of the variance in multitasking performance, visual attention control uniquely explained 13.2% of the variance, and auditory and visual attention control shared 30.3% of the variance in multitasking performance (see Fig. 18). Thus, auditory attention control contributed little to the prediction of multitasking performance above and beyond visual attention control; however, auditory and visual attention control shared a substantial proportion of the variance that they accounted for.

Discussion

To investigate the extent to which attention control is a modality-general construct, we developed three auditory tests of attention control based on classic conflict paradigms and incorporated an adaptive response deadline

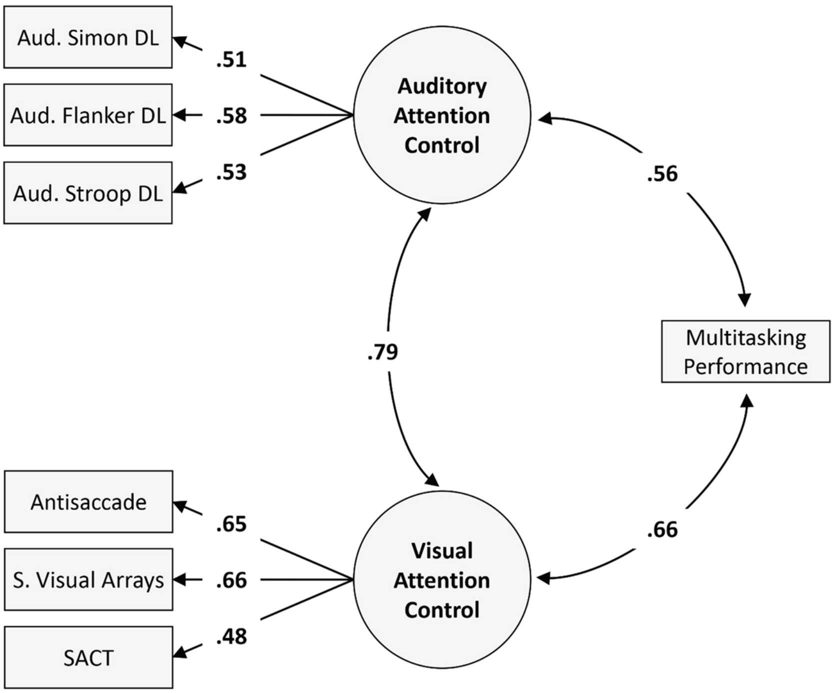

Fig. 16 Confirmatory factor analysis depicting correlations between an auditory attention control factor, a visual attention control factor, and multitasking performance. χ^2 (12) = 20.33, $p = .061$; CFI = .976, TLI = .958, RMSEA = .046, 90% CI [.000, .080], SRMR = .032

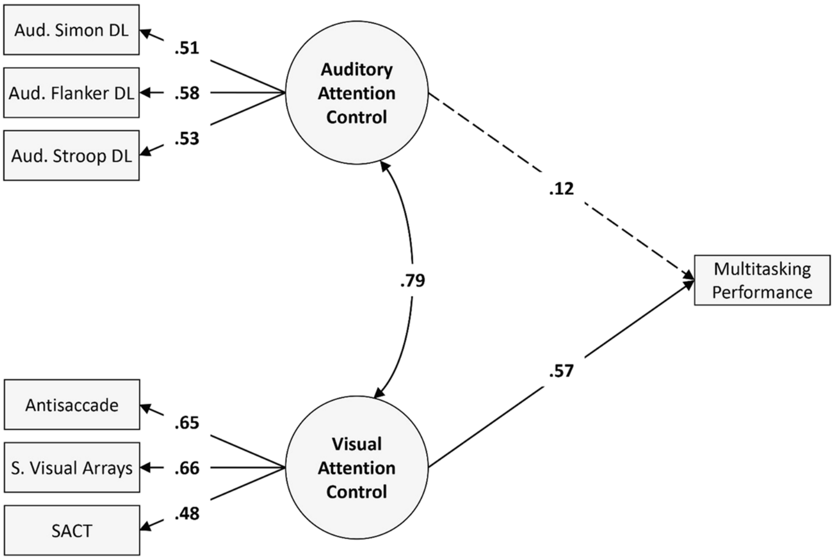

Fig. 17 Structural equation model with an auditory attention control factor and a visual attention control factor predicting multitasking performance. χ^2 (12) = 20.33, $p = .061$; CFI = .976, TLI = .958, RMSEA = .046, 90% CI [.000, .080], SRMR = .032

(DL) to account for speed–accuracy trade-offs. We named the tasks Auditory Simon DL, Auditory Flanker DL, and Auditory Stroop DL. From a psychometric standpoint, the three auditory conflict tasks were reliable and efficient. Estimates of internal consistency ranged from α = .89 to .92 for the final four reversals of the adaptive staircase, correlations between the response deadlines at the midpoint and conclusion of the task ranged from $r = .97$ to .99, and mean testing time ranging from 5 to 8 minutes.

Critical to the success of the three auditory conflict tasks was the degree to which the staircase procedure, which adjusted the duration of the response deadline depending on participants' performance, effectively converged on the desired accuracy rate of 75% on incongruent trials. According to Kaernbach (1991),

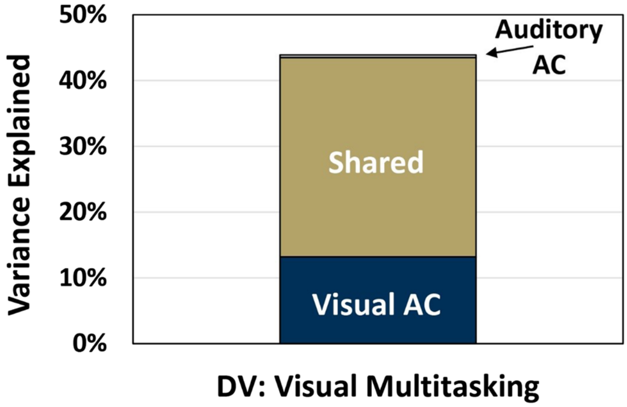

Fig. 18 Commonality analysis decomposing variance in visual multitasking performance

the process should converge on 75% accuracy because there were two response options on each trial, and the step-size ratio for incorrect versus correct responses was set to 3:1. Indeed, averaging across participants, mean accuracy rates on incongruent trials were within ±1% of 75%, providing strong evidence that the staircase procedure worked as intended.

We highlight the importance of the adaptive response deadline procedure in these tasks because it provides a method of handling speed–accuracy trade-offs, which are known to pose problems for studies of individual differences in attention control (Draheim et al., 2019; Heitz, 2014). The three auditory conflict tasks account for speed–accuracy trade-offs by holding the accuracy rate on incongruent trials constant at 75%. With accuracy constrained to be equal across participants, the tasks use differences in the speed with which subjects can respond with 75% accuracy as the measure of performance. Specifically, the outcome measure is computed as the average response deadline over the final four reversals of the staircase procedure.[3] While our approach of using an adaptive response deadline is not the only way to handle speed–accuracy trade-offs, the results following from these tasks seem to indicate that it provides a viable method for addressing the issue.

Having established that the adaptive staircase procedure worked as intended, we investigated the modality-specificity of attention control using a factor-analytic approach. Specifically, we compared the three auditory conflict tasks to visual attention control tests as well as tests of fluid intelligence, working memory capacity, processing speed, dichotic listening, and multitasking performance. An exploratory factor analysis provided the first evidence of modality-generality; the auditory and visual attention control tests loaded highly on the first extracted factor, and the auditory attention control tests had low cross-loadings on the other extracted factors. Using confirmatory factor analysis, we found that the auditory and visual attention control tests loaded well on a common factor, demonstrating good fit and strong, balanced factor loadings. That said, a two-factor model fit the data slightly better, indicating that the auditory attention control tasks and visual attention control tasks were better modeled as loading on separate but highly correlated factors ($r = .81$). This provides evidence for modality-specific or method-specific effects in the measurement of attention control. Nevertheless, further confirmatory factor analyses revealed that auditory attention control correlated significantly more strongly with visual attention control than with fluid intelligence, working memory capacity, and processing speed. Thus, auditory attention control was more strongly related to visual attention control than to the other cognitive ability factors.

We tested whether processing speed could account for the correlation between auditory and visual attention control factors, because both sets of tasks require speeded perceptual encoding and decision-making, and, for the auditory tasks, speeded motor execution. We used a latent perceptual processing speed factor to control for these sources of variance, because the processing speed tasks required speeded perceptual encoding, decision-making, and response execution. Although a latent factor representing perceptual processing speed significantly predicted auditory and visual attention control, after controlling for processing speed, the correlation between visual and auditory attention control factors was only reduced from $r = .81$ to $r = .75$ (i.e., a nonsignificant reduction). Thus, the vast majority of the covariance between the auditory and visual attention control factors does not seem attributable to the speed of perceptual processing, providing support for the validity of the new auditory attention control measures. Nevertheless, our operationalization of perceptual speed was based on tasks involving rapid comparisons of relatively simple stimuli (e.g., letters, digits, and patterns). These results could be extended by considering a larger taxonomy of tasks or measures reflecting "processing speed" (e.g., Sheppard & Vernon, 2008).

In our final set of analyses, we tested whether the unique variance that distinguished auditory attention control from visual attention control contributed to dichotic listening performance or multitasking performance. Indeed, auditory and visual attention control were both highly correlated with dichotic listening performance ($r = .57$ vs. $r = .50$), but only the unique contribution of auditory attention control was statistically significant ($\beta = .47$, $p = .038$, vs. $\beta = .12$, $p = .605$). Auditory attention control uniquely explained 8.5% of the variance in dichotic listening, visual attention control uniquely explained 1.2% of the variance, and auditory and visual attention control shared 23.1% of the variance in dichotic listening performance.

---

[3] The tasks are available for download on the Open Science Framework, along with the raw data and R code used to score the data files and generate figures (https://osf.io/2zqe7/).

By contrast, though auditory and visual attention control were highly correlated with multitasking performance ($r = .56$ vs. $r = .66$), only the unique contribution of visual attention control was statistically significant ($\beta = .12$, $p = .486$ vs. $\beta = .57$, $p < .001$). Auditory attention control uniquely explained 0.4% of the variance in multitasking performance, visual attention control uniquely explained 13.2% of the variance, and auditory and visual attention control shared 30.3% of the variance in multitasking performance. Importantly, the majority of the variance that the auditory and visual attention control factors accounted for in dichotic listening and multitasking performance was shared across modalities, suggesting an important role for domain-general attention control.

It is perhaps unsurprising that the auditory attention control factor was not perfectly isomorphic with the visual attention control factor; this difference could be attributed to method-specific effects (e.g., the use of an adaptive deadline procedure for the auditory tasks but not for the visual tasks) and/or modality-specific effects (e.g., the use of auditory stimuli vs. visual stimuli). However, the incremental validity of auditory attention control for predicting dichotic listening suggests that part of what differentiated auditory from visual measures of attention control is not merely the use of an adaptive response deadline procedure. As an auditory-modality task, what dichotic listening uniquely shares with the auditory attention control measures but not with visual attention control measures is *modality-specific variance*. Similarly, it can be seen as a double dissociation that the measures of auditory attention control *did not* contribute incrementally to the prediction of multitasking performance above and beyond visual attention control, because the multitask used in the present study relied on the visual modality. Taken together, these results suggest that modality-specific variance is shared across distinct tests of cognitive ability. We urge researchers to follow up on these results, and offer the three auditory conflict tasks as tools to facilitate that work.

Implications for the modality-specificity of attention control

The results of the present study can be interpreted from a neuroscientific perspective to better understand why auditory and visual attention control measures demonstrated evidence for both unity and diversity. Importantly, differences in performance across modalities does not provide strong evidence for modality-specificity of attention control, because differences could emerge very early on in the stream of information processing, before attention necessarily plays a role. Performance can therefore be seen as reflecting the interplay between domain-general attention control and influences that are particular to the specific processing sites for different sensory modalities.

Our senses are processed through diverging pathways in the brain; visual information from the retina is sent to the lateral geniculate nucleus in the thalamus prior to the primary visual cortex (i.e., V1) (Kandel et al., 2000), whereas auditory information enters the cochlea and ascends to the medial geniculate before making its way to the primary auditory cortex (i.e., A1) (Purves & Williams, 2001). Damage to the retina (or alternatively, the cochlea) could lead to differences in performance on visual versus auditory tests of attention control, not because performance in these sensory modalities is supported by different attentional systems, but simply because there might be problems affecting early sensory processing in one modality and not the other (Jan et al., 2019).

The ongoing debate is whether there is a modality-general (i.e., "supramodal") *source of attention control* that exerts top-down influence through descending connections on the processing sites of different modalities. Neuroimaging studies suggest that the frontoparietal control network is a candidate network for attention control; regions identified as part of this network demonstrate significant activity during both auditory and visual conflict tasks (Dosenbach et al., 2006; Fan, 2014; Green et al., 2011; Spagna et al., 2015; Wu et al., 2020). Furthermore, Kerlin et al. (2010) found overlapping alpha modulation (8 –12 Hz) at the parietal-occipital sites across hemispheres in a task that mimics the "cocktail party effect." Crucially, the degree of lateralization when listening to speech, which was used as an indication of the spatial direction of auditory attention, predicted an individual's attentional gain and performance on a speech-in-noise task (Kerlin et al., 2010).

Posner and Driver (1992) define the primary sensory cortices (e.g., V1 and A1) as the earliest sites at which attention control exerts its influence. Using magnetoencephalography (MEG), Poghosyan and Ioannides (2008) found that during a visual and auditory detection task, attention modulated A1 as early as 30–50 ms and V1 as early as 55–90 ms. Furthermore, patients with parietal lesions have been shown to struggle to use visual and auditory cues in a spatial attention task (Farah et al., 1989). These interactions have been used as evidence for a supramodal view of attention control.

Nevertheless, if attention control is unitary across modalities, we must explain why the two-factor model fit the data slightly better than the single-factor model, because this could be used to argue for the modality-specificity of attentional processes. Indeed, some researchers have advocated for this view: Lin et al. (2017) found that attentional deficits in attention-deficit/hyperactivity disorder (ADHD) were modality-specific, concluding that visual deficits were more serious than those of auditory deficits in their sample.

An alternative explanation for this observation is that higher- or latent-level processes such as attention control are domain-general while lower-order processes may have modality-specific influences. Using the Attention Network Test (Fan et al., 2002), Spagna et al. (2015) found that executive control, but not alerting or orienting, was significantly correlated across visual and auditory tasks. They suggested

Table 5 Performance on congruent and incongruent trials in the adaptive response deadline auditory conflict tasks

Measures	*M*	*SD*	*r*	Paired-samples *t*-test	Cohen's *d*
Stroop congruent ACC	0.82	0.07	.46	$t(309) = 18.91, p < .001$	1.074
Stroop incongruent ACC	0.76	0.04	--	--	--
Stroop congruent RT	748.60	80.35	.98	$t(309) = -17.44, p < .001$	−0.99
Stroop incongruent RT	766.32	83.45	--	--	--
Flanker congruent ACC	0.78	0.08	.67	$t(300) = 9.01, p < .001$	0.52
Flanker incongruent ACC	0.75	0.07	--	--	--
Flanker congruent RT	681.11	107.12	.97	$t(300) = -4.02, p < .001$	−0.23
Flanker incongruent RT	687.67	117.56	--	--	--
Simon congruent ACC	0.82	0.08	.50	$t(299) = 17.62, p < .001$	1.02
Simon incongruent ACC	0.74	0.06	--	--	--
Simon congruent RT	538.49	105.43	.98	$t(299) = -13.44, p < .001$	−0.78
Simon incongruent RT	557.84	120.66	--	--	--

ACC = accuracy, RT = response time. Note that performance on incongruent and congruent trials is entwined through the structure of the task and the use of an adaptive response deadline conditioned on incongruent trials

that executive control of attention may be modality-general, while alerting and orienting functions may rely more upon modality-specific processes. That said, the reliability of the Attention Network Test is questionable (Ishigami et al., 2016) because the measures of performance are based on difference scores. Therefore, conclusions based on the strength of these relationships should be interpreted with caution. As others have suggested (Corbetta, 1998; Driver & Frackowiak, 2001; Miller & Cohen, 2001; Spagna et al., 2015), attentional effects might reflect the interplay between a domain-general attention control system and the workings of modality-specific processing sites.

Limitations

In future work, we plan to assess the test–retest reliability of the three auditory DL tests of attention control to ascertain whether the measures capture trait-level stability in executive functioning. Additionally, it would be illuminating to assess state-level changes in performance as a function of, for instance, cognitive load or sleep deprivation. Such work could shed light on more fundamental questions related to the nature and measurement of attention control, such as whether the same factors that negatively impact performance on visually-based attention control tests also impact performance on auditory-based attention control tests.

Conclusion

We investigated modality-specific effects on attention control by developing three efficient and reliable tests of auditory attention control that use an adaptive response deadline to account for speed–accuracy trade-offs. The results provide evidence for a common attentional ability underpinning performance on attention-demanding tests in the visual and auditory modalities. By making the three auditory conflict tasks (i.e., Auditory Simon DL, Auditory Flanker DL, and Auditory Stroop DL) freely available to researchers, future work will be better able to illuminate sources of convergence and divergence in cognitive performance across sensory modalities.

Appendix

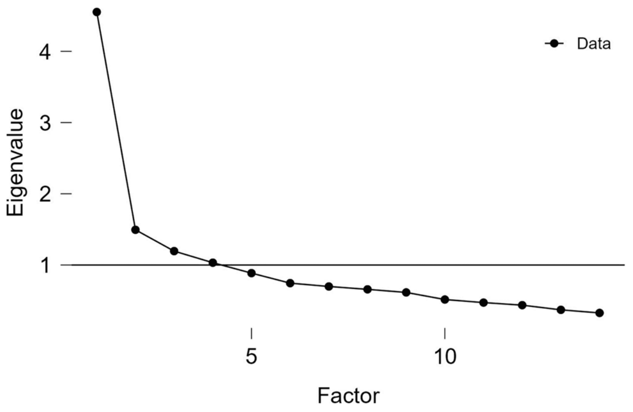

Fig. 19 Scree plot for exploratory factor analysis reported in Table 4

Author Note Data, task downloads, and R code are openly available at https://osf.io/2zqe7/.

Funding This work was supported by Office of Naval Research grants N00014-21-1-2327 and N00014-17-1-2061 and Naval Research Lab grants N00173-20-2-C003 and SA-GAT-NRL-0135-2019 to Randall W. Engle, and Office of Naval Research grant N00014-22-S-F002 to Randall W. Engle and Alexander P. Burgoyne.

References

Anderson, J. R. (2005). *Cognitive psychology and its implications*. Macmillan.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 5–28.

Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science, 29*(6), 624–630.

Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., & Engle, R. W. (2022). Attention control and process overlap theory: Searching for cognitive processes underpinning the positive manifold. *Intelligence, 91*, 101629.

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General, 152*(8), 2369–2402. https://doi.org/10.1037/xge0001408

Chan, J. S., Merrifield, K., & Spence, C. (2005). Auditory spatial attention assessed in a flanker interference task. *Acta acustica united with acustica, 91*(3), 554–563.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975–979.

Christensen, T. A., Lockwood, J. L., Almryde, K. R., & Plante, E. (2011). Neural substrates of attentive listening assessed with a novel auditory Stroop task. *Frontiers in Human Neuroscience, 4*, 236.

Conners, C. K. (2014). *Conners continuous auditory test of attention*. (Conners CATA).

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8*, 331–335.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences, 95*(3), 831–838.

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron, 50*(5), 799–812.

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment, 30*(1), 116.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin, 145*(5), 508.

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General, 150*(2), 242.

Draheim, C., Tshukara, J. S., & Engle, R. W. (2023). Replication and extension of the toolbox approach to measuring attention control. *Behavior Research Methods*, 1–23.

Driver, J., & Frackowiak, R. S. (2001). Neurobiological measures of human selective attention. *Neuropsychologia, 39*(12), 1257–1262.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests: 1976*. Educational Testing Service.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*(2), 190–193.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(*1*), 143–149. https://doi.org/10.3758/BF03203267

Fan, J. (2014). An information theory account of cognitive control. *Frontiers in human neuroscience, 8*, 680.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347.

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia, 27*(4), 461–470.

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General, 129*(1), 61.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences, 16*(2), 129–135.

Green, J. J., Doesburg, S. M., Ward, L. M., & McDonald, J. J. (2011). Electrical neuroimaging of voluntary audiospatial attention: Evidence for a supramodal attention control network. *Journal of neuroscience, 31*(10), 3560–3564.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18*, 1279–1296.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166–1186.

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*, 150.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 645–662. https://doi.org/10.1037/0278-7393.33.4.645

Ishigami, Y., Eskes, G. A., Tyndall, A. V., Longman, R. S., Drogos, L. L., & Poulin, M. J. (2016). The Attention Network Test-Interaction (ANT-I): Reliability and validity in healthy older adults. *Experimental Brain Research, 234*, 815–827.

Jan, C., Li, S., Kang, M., Liu, L., He, L., Jin, L., Qin, X., Congdon, N., & Wang, K. (2019). Association of visual acuity with educational outcomes: A prospective cohort study. *British Journal of Ophthalmology, 103*(*11*), 1666–1671. https://doi.org/10.1136/bjophthalmol-2018-313294

JASP Team (2023). JASP (Version 0.17.1) [Computer software].

James, W. (1890). *The Principles of Psychology, I.*

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49*(3), 227–229.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (4th ed.). McGraw-Hill.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity.

Journal of Experimental Psychology: General, 130(2), 169–183. https://doi.org/10.1037/0096-3445.130.2.169

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience, 30*(2), 620–628.

Kline, R. B. (2023). *Principles and practice of structural equation modeling. Guilford* publications.

Lin, H. Y., Hsieh, H. C., Lee, P., Hong, F. Y., Chang, W. D., & Liu, K. C. (2017). Auditory and visual attention performance in children with ADHD: The attentional deficiency of ADHD is modality specific. *Journal of Attention Disorders, 21*(10), 856–864.

Martin, J., Mashburn, C. A., & Engle, R. W. (2020). Improving the Validity of the Armed Service Vocational Aptitude Battery with Measures of Attention Control. *Journal of Applied Research in Memory and Cognition, 9*(3), 323–335.

Martin, J. D., Tsukahara, J. S., Draheim, C., Shipstead, Z., Mashburn, C. A., Vogel, E. K., & Engle, R. W. (2021). The visual arrays task: Visual storage capacity or attention control? *Journal of Experimental Psychology: General, 150*(12), 2525.

Mashburn, C. A., Burgoyne, A. P., & Engle, R. W. (2023). *Working memory, intelligence, and life success* (p. 149). *Memory in Science for Society*.

Mashburn, C. A., Burgoyne, A. P., & Engle, R. W. (under review). Knowledge, attention, and psychomotor ability: A latent variable approach to understanding individual differences in simulated work performance.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167–202.

Nimon, K., Henson, R. K., & Gates, M. S. (2010). Revisiting interpretation of canonical correlation analysis: A tutorial and demonstration of canonical commonality analysis. *Multivariate Behavioral Research, 45*(4), 702–724.

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin, 82*(1), 85.

Poghosyan, V., & Ioannides, A. A. (2008). Attention modulates earliest responses in the primary auditory and visual cortices. *Neuron, 58*(5), 802–813.

Posner, M. I., & Driver, J. (1992). The neurobiology of selective attention. *Current Opinion in Neurobiology, 2*(2), 165–169.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/

Purves, D., & Williams, S. M. (2001). *Neuroscience* (2nd ed.). Sinauer Associates.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales* (Vol. 759). Oxford Psychologists Press.

Redick, T. S., Heitz, R. P., & Engle, R. W. (2007). Working memory capacity and inhibition: Cognitive and social consequences. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 125–142). American Psychological Association. https://doi.org/10.1037/11587-007

Redick, T. S., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology, 24*(7), 844–854.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review, 26*(2), 452–467.

Rouder, J. N., Kumar, A., & Haaf, J. M. (2023). Why Many Studies of Individual Differences With Inhibition Tasks May Not Localize Correlation. *PsyArXiV Preprint.* https://doi.org/10.31234/osf.io/fsqvu

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*(5), 763.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of experimental psychology: General, 125*(1), 4.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and individual differences, 44*(3), 535–551.

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*(3), 300–304. https://doi.org/10.1037/h0020586

Sohn, Y. W., & Doane, S. M. (2003). Roles of working memory capacity and long-term working memory skill in complex task performance. *Memory & Cognition, 31*, 458–466.

Spagna, A., Mackie, M. A., & Fan, J. (2015). Supramodal executive control of attention. *Frontiers in Psychology, 6*, 65.

Spearman, C. (1904). " General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201–292.

Sperry, R. W. (1993). The impact and promise of the cognitive revolution. *American Psychologist, 48*(8), 878.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Thurstone, L. L. (1938). *Primary mental abilities. Psychometric Monographs.*

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior, 3*(6), 449–459.

Tsukahara, J. S., & Engle, R. W. under review. *Sustaining the focus of attention and how it relates to performance in complex cognitive tasks.* https://osf.io/wd5kz/download

Tsukahara, J. S., Harrison, T. L., Draheim, C., Martin, J. D., & Engle, R. W. (2020). Attention control: The missing link between sensory discrimination and intelligence. *Attention, Perception, & Psychophysics, 82*, 3445–3478.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1302.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438(7067), 500–503.*

Walker, P. B., Olde, B. A., & Olson, T. M. (2007). *Improving Aviator Selection Using the Performance-Based Measurement Battery (PBMB). Poster presented at the Proceedings of the Human Systems Integration Symposium.* Annapolis Retrieved from https://www.researchgate.net/publication/266910869_Improving_Aviator_Selection_Using_the_Performance-Based_Measurement_Battery_PBMB

Wang, L. J., Lee, S. Y., Tsai, C. S., Lee, M. J., Chou, M. C., Kuo, H. C., & Chou, W. J. (2021). Validity of visual and auditory attention tests for detecting ADHD. *Journal of Attention Disorders, 25*(8), 1160–1169.

Wood, N. L., & Cowan, N. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General, 124*(3), 243.

Wu, T., Chen, C., Spagna, A., Wu, X., Mackie, M. A., Russell-Giller, S., et al. (2020). The functional anatomy of cognitive control: A domain-general brain network for uncertainty processing. *Journal of comparative neurology, 528*(8), 1265–1292.