Auditory Attentional Load Modulates Audiovisual Integration During Auditory/Visual Discrimination

Yanna Ren[1], Nengwu Zhao[1], Junyuan Li[1], Junhao Bi[1], Tao Wang[3], and Weiping Yang[2]*

[1] Department of Psychology, College of Humanities and Management, Guizhou University of Traditional Chinese Medicine, Guiyang, 550025, China.

[2] Department of Psychology, Faculty of Education, Hubei University, Wuhan, 430062, China.

3 Department of Light and Chemical Engineering, Guizhou Light Industry Technical College, Guiyang, 550025, China.

KEYWORDS

audiovisual integration,
attentional load,
auditory attention,
discrimination task,
race model

ABSTRACT

Attention modulates numerous stages of audiovisual integration, and studies have shown that audiovisual integration is higher in attended conditions than in unattended conditions. However, attentional resources are limited for each person, and it is not yet clear how audiovisual integration changes under different attentional loads. Here, we explored how auditory attentional load affects audiovisual integration by applying an auditory/visual discrimination task to evaluate audiovisual integration and a rapid serial auditory presentation (RSAP) task to manipulate auditory attentional resources. The results for peak benefit and positive area under the curve of different probability showed that audiovisual integration was highest in the low attentional load condition and lowest in the high attentional load condition (low > no = medium > high). The peak latency and time window revealed that audiovisual integration was delayed as the attentional load increased (no < low < medium < high). Additionally, audiovisual depression was found in the no, medium, and high attentional load conditions but not in the low attentional load condition. These results suggest that mild auditory attentional load increases audiovisual integration, and high auditory attentional load decreases audiovisual integration.

INTRODUCTION

Individuals are often inundated with stimuli from various sensory modalities (e.g., auditory, visual, tactile, olfactory, and somatosensory stimuli). However, our brains can effectively screen out and integrate the most pertinent information to better understand our surroundings. For example, to safely drive a car, we must comprehensively utilize auditory information (e.g., music, car engine), visual information (e.g., road, roadside billboards), and tactile information (e.g., feeling the steering wheel). The integration of information from various sensory modalities is called multisensory integration (Meredith et al., 1987; Stein & Meredith, 1993). In the real world, auditory and visual modalities are common forms of acquiring sensory information, and the investigation for integration of auditory and visual information is the focus in multisensory integration. Studies have confirmed that

responses to bimodal audiovisual stimuli are faster than responses to unimodal auditory or visual stimuli (Meredith et al., 1987; Stein & Meredith, 1993). For example, movies with sound are considerably easier to understand than silent films and audio shows. Furthermore, attention is a key factor that alters the processing of sensory stimuli by facilitating the detection and perception of the stimulus that is presented at the attended location (Ho et al., 2009; Mcdonald et al., 2000). Numerous studies have confirmed that attention affects audiovisual integration in multiple stages and that audiovisual integration is higher

Corresponding author: Weiping Yang, Department of Psychology, Faculty of Education, Hubei University, Wuhan, 430062, China.

Email: swywp@163.com

for attended conditions than for unattended conditions (Talsma et al., 2007; Talsma et al., 2010; Talsma et al., 2009; Talsma & Woldorff, 2005; Xu et al., 2020).

According to the attentional load theory proposed by Lavie (1995) and Lavie and Tsal (1994), attentional resources are limited for each person, and when multiple tasks are conducted simultaneously, if one task occupies more attentional resources, less attention will be allocated to the other tasks. In everyday life, individuals continuously face more than one event simultaneously, and the main task is always disturbed by distractors in the surrounding environment (Fan et al., 2002; Stein, 2012). There is little research on whether our brain can integrate the effective auditory and visual information by suppressing other distractors during these situations. To clarify this matter, Alsius et al. (2014; Alsius et al., 2005) examined audiovisual integration using the classical McGurk effect under low or high visual attentional load conditions. In the low visual attentional load, the participants were instructed to conduct a single McGurk task, but in the high visual attentional load condition, they were instructed to conduct dual tasks (a McGurk task and a rapid serial visual presentation, RSVP, task). Both the behavioral (Alsius et al., 2005) and event-related potential (Alsius et al., 2014) results showed that the audiovisual integration was weaker under the high visual attentional load condition than under the low visual attentional load condition. In view of the limitation of attentional resources (Lavie, 1995; Lavie & Tsal, 1994), the most likely explanation was that fewer attentional resources were left to process the McGurk words under the high visual attentional load condition than under the low visual attentional load condition because the demand for attentional resources by the distractor stimuli was larger under the high-load condition (Alsius et al., 2014; Alsius et al., 2005). Alsius et al. further concluded that attentional load decreases the audiovisual integration. As the audiovisual integration was larger when the stimuli were presented at the attended location than when presented at the unattended location (Talsma et al., 2007; Talsma et al., 2010; Talsma et al., 2009; Talsma & Woldorff, 2005), audiovisual integration might increase with the addition of attentional resources to process the audiovisual information. However, in the studies by Alsius et al. (2005, Alsius et al., 2014) there were only two attentional load conditions (low and high), and the attentional load was too high in the high load condition, yielding an accuracy of less than 60% in the dual tasks (Alsius et al., 2005; Alsius et al., 2014). It cannot be determined whether the audiovisual integration was reduced due to the direct addition of attentional load or whether it increased under low attentional load and then reduced inversely under high attentional load. Additionally, speech materials were employed in the studies by Alsius et al., which included high-order speech processing. Therefore, it is difficult to determine whether the attentional load influenced the audiovisual integration or whether it influenced the high-order speech processing.

By removing the effect of high-order cognitive processing, Ren, Li et al. (2020) and Ren, Zhou et al. (2020) investigated how the audiovisual integration changes with increasing visual attentional load using a meaningless auditory/visual discrimination task and an RSVP task. Similar to Alsius et al.'s (2005, Alsius et al., 2014) experimental design,

single (no and low loads) or dual tasks (high load) were performed. Their results showed that the audiovisual integration was higher in the low attentional load condition than in the no or high attentional load conditions. Together with Alsius et al.'s studies, Ren, Li et al. (2020) and Ren, Zhou et al. (2020) proposed that low visual attentional load increases audiovisual integration, and high visual attentional load decreases audiovisual integration. Notably, the aforementioned studies employed transient distractors that induced transient attention. In contrast to transient attention, in which the participant is occasionally cued to the stimulated location, sustained attention maintained over time on a specific task affects the information perception differently (Eimer, 1996; Eimer & Forster, 2003). To investigate the influence of sustained attention on audiovisual integration, Wahn and König (2015) instructed participants to continuously track visual moving balls when performing an audiovisual redundancy task. The results revealed that audiovisual integration was comparable under both the high and low sustained perceptual load conditions, indicating that sustained visual attentional load does not significantly affect audiovisual integration. Wahn and König (2017) reported that shared or distinct attentional resources across sensory modalities are task dependent, where auditory and visual attentional resources are distinct during the discrimination of stimulus attributes (Alais et al., 2006; Arrighi et al., 2011) but are shared during stimulus location (Driver & Spence, 1998a, 1998b; Spence, 2010a, 2010b; Spence & Driver, 2004). In studies by Ren, Li, et al. (2020), Alsius et al. (2005; Alsius et al., 2014), and Wahn et al. (2017; Wahn & König, 2015), the auditory/visual discrimination task involved the discrimination of stimulus attributes, and the second distractor task was from a visual modality. However, whether auditory attentional load influences audiovisual integration has not been determined.

Furthermore, a recent study illustrated that the integration of auditory and visual information is rearranged according to the auditory signal during audiovisual integration (Spence & Squire, 2003). The McGurk effect clearly displays visual dominance where what is being heard is influenced by what is being seen (McGurk & MacDonald, 1976). However, the modulation of auditory information on visual perception has also been extensively reported (Morein-Zamir et al., 2003; Sekuler et al., 1997; Shams et al., 2002; Stein et al., 1996). Shams et al. (2002) instructed participants to judge the number of flashes presented on a monitor screen. The participants reported two or more flashes when one flash was accompanied by two or more beeps, exhibiting the phenomenon of sound-induced illusory flashing, which suggested that the auditory beep altered visual perception. Consistent results were also obtained by Sekuler et al. (1997) using a temporal order judgment task (Morein-Zamir et al., 2003). These results indicate that perceptual reports tend to be biased to auditory signals when an audiovisual stimuli pair is presented. Therefore, it is taken for granted that auditory attentional load affects audiovisual integration. Additionally, to fully understand the interplay between attention and audiovisual integration, it is necessary to investigate audiovisual integration under different auditory attentional load conditions. Therefore, the interest of the present study was to clarify how auditory attentional load influences

audiovisual integration. Attentional load theory proposes that if all tasks are easy enough, they can be completed successfully. However, if the tasks are difficult, the main task can be completed by allocating more attentional resources, while the other tasks cannot be completed because of attentional resource exhaustion (Lavie, 1995; Lavie & Tsal, 1994). In addition, attention can facilitate the detection and perception of stimuli that are presented at attentional locations (Petersen & Posner, 2012; Posner & Petersen, 1990), and the audiovisual integration is higher at the attended location than at the unattended location (Talsma et al., 2007; Talsma et al., 2010; Talsma et al., 2009; Talsma & Woldorff, 2005). Adaptive stress can produce optimal performance (Kamijo et al., 2007; Yerkes & Dodson, 1908). Therefore, we hypothesized that audiovisual integration would be higher in a low attentional load condition than in a no attentional load condition. However, according to the perceptual load theory (Lavie, 1995; Lavie & Tsal, 1994), more attentional resources might be delivered to the secondary distractors as the attentional load increases, and less attentional resources would be left to process the auditory/visual discrimination task. Therefore, we further hypothesized that audiovisual integration would decrease inversely in the high auditory attentional load condition.

METHODS

Subjects

Twenty young healthy students (18–21 years, $M_{age} \pm SD$, 19.44 ± 1.00) participated in this study, and all student volunteers were recruited from Guizhou University of Traditional Chinese Medicine. All participants were free of neurological diseases, had normal hearing and normal or corrected-to-normal vision, had no color blindness or color weakness, and were naive to the purpose of the experiment. Additionally, all participants were paid for their time and provided written informed consent to participate, which was previously approved by the ethics committee of the Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine (No. 2018072).

Stimuli and Task

The participants were instructed to perform single and dual tasks. These were an auditory/visual discrimination task to evaluate audiovisual integration and an RSAP task to manipulate the attentional load by competing attentional resources away from the auditory/visual discrimination task. According to the attentional load session, the auditory/visual discrimination task and RSAP task were presented simultaneously or separately.

For the auditory/visual discrimination task, the visual nontarget stimulus was a black and white checkerboard image (B/W checkerboard, 52 × 52 mm, with a visual angle of 5 °), and the visual target stimulus was a black-and-white checkerboard image with two black dots contained within each white checkerboard (He et al., 1996; Laura et al., 2005; Ren et al., 2016; see Figure 1, Panel A). The auditory nontarget stimulus was a 1000-Hz sinusoidal tone, and the auditory target stimulus was white noise (Ren et al., 2018; Ren et al., 2016; Yang et al.,
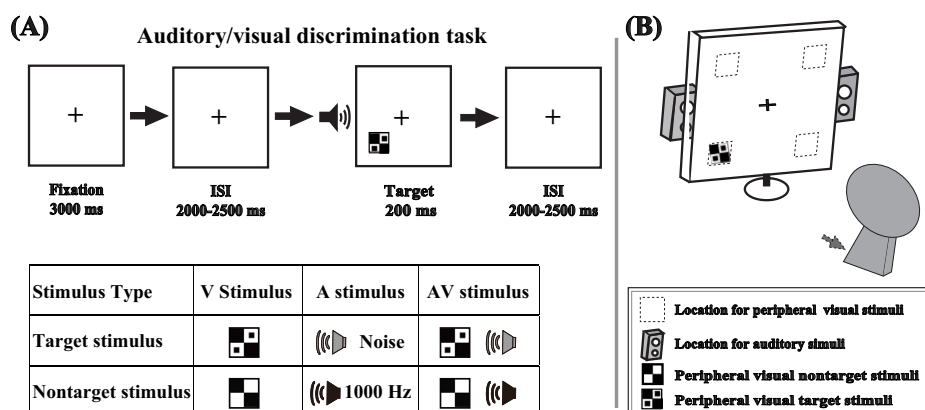
2015). In line with previous studies about the effect of visual perceptual load on audiovisual integration (Ren, Li, et al., 2020; Ren, Zhou, et al., 2020), the visual stimuli (V) were presented on a computer monitor in front of participants' eyes and on the upper/lower left or right quadrant of the screen for 200 ms with a 12 ° visual angle (see Figure 1, Panel B, gray square). The auditory stimuli (A) were presented through two speakers at approximately 60 dB SPL for a duration of 200 ms (10 ms of the rise/fall cosine gate). The audiovisual target stimulus (AV) was the combination of the visual target and auditory target stimuli, and the audiovisual nontarget stimulus was the combination of the visual nontarget and auditory nontarget stimuli. The following conditions were not included: a visual target stimulus accompanied by an auditory nontarget stimulus and a visual nontarget stimulus accompanied by an auditory target stimulus.

The stimuli in the RSAP task consisted of 10 distractor characters taken from six letters (B, C, P, R, T, and V) and four digits (6, 7, 8, and 9) presented through the speakers located on the right/left sides of the computer monitor (see Figure 1, Panel B). The stimuli of the RSAP were compounded with the auditory stimuli of the auditory/visual discrimination task using Audacity 2.4.2 (https://www.audacityteam.org/).

Procedure

Participants were instructed to perform the experiment in a dimly lit and sound-attenuated room (neuroscience laboratory room, Guizhou University of Traditional Chinese Medicine, China). Four separate sessions were conducted, including a no attentional load session, a low attentional load session, a medium attentional load session, and a high attentional load session. The order of the four sessions was counterbalanced across participants.

In the no attentional load session, a single task was applied. A fixation cross was presented for 3000 ms, and the stimuli of the auditory/visual discrimination task were subsequently presented randomly with a random interstimulus interval (ISI) from 2000 ms to 2500 ms (see Figure 1, Panel A). Because there was no other additional distractor to compete for attentional resources, it was called the "no attentional load session." In the no attentional load session, there were 80 trials for each target stimulus type (A, V, and AV) and 20 trials for each nontarget stimulus type (A, V, and AV), with an appropriate break according to the specific situation of each participant. Participants were instructed to press the left mouse button to respond to the target stimuli as rapidly and as accurately as possible. In the low attentional load session, the same single task was performed as in the no attentional load session. However, differently, the RSAP task was presented simultaneously through the speakers located on the right/left side of the monitor, with an identical probability for each letter. Although the participants were instructed to respond to the target stimuli of the auditory/visual discrimination task only, it was impossible for the participants to completely neglect the distractor. Therefore, it was called the "low attentional load session." According to previous studies, object recognition is more difficult than object classification (Goldhammer et al., 2014; Houlihan et al., 1998). Therefore, a dual task was performed in the medium and high attentional load sessions such that attentional resources

FIGURE 1.

Schematic depiction and the stimuli used in the auditory/visual discrimination task (Panel A). Audiovisual integration was evaluated using an auditory/visual discrimination task, which was presented peripherally (gray square and speakers), and attentional load was manipulated using an RSAP task, which was presented through a speaker located on the right/left side of the monitor (Panel B).

would have to compete with the auditory/visual discrimination task. In the medium attentional load condition, the presentation for both the auditory/visual discrimination task and the RSAP task were identical to that of the low attentional load session, but the participants were instructed to press the left mouse button to respond to the target stimuli of the auditory/visual discrimination task and to press the right mouse button to respond to the digits (6, 7, 8, and 9) of the RSAP task as rapidly and accurately as possible (object classification). In the high attentional load condition, the presentations of both the auditory/visual discrimination task and the RSAP task were identical to those of the low attentional load session, but the participants were instructed to press the left mouse button to respond to the target stimuli of the auditory/visual discrimination task and the right mouse button to respond to "8" and "B" of the RSAP task as rapidly and accurately as possible (object recognition). During the medium and high attentional load sessions, the participants were instructed to treat the auditory/visual discrimination task and RSAP task equally.

Data Analysis

The hit rate is the percentage of correct responses (the response time falling within the average time duration ± 2.5 SD) relative to the total number of target stimuli. The hit rates and response times (RTs) were computed separately for each participant under each condition, and the data were then subjected to a 4 (Attentional load: no, low, medium, and high) × 3 (Target modality: A, V, and AV) analysis of variance (ANOVA, Greenhouse-Geisser corrections with corrected degrees of freedom). The statistical significance level was set at $p < .05$, and the effect size estimates, η_p^2, are also reported.

Additionally, to assess the effect of attentional load on audiovisual integration, the amount of audiovisual integration was calculated using a race model based on cumulative distribution functions (CDFs) of the response time of each trial (Miller, 1982, 1986). The independent race model is a statistical prediction model based on the CDFs of the summed probabilities of the visual and auditory responses to inde-

pendent unimodal visual and auditory stimuli. This model allows for a direct comparison of the probability of the multisensory condition to the predicted probability of the unimodal conditions, [P(V)+P(A)-P(V)×P(A)], by segmenting the participant-specific CDFs for each condition using 10 ms time bins (Miller, 1982, 1986). P(V) is the probability of responding within a given time in a unimodal visual trial and P(A) is the probability of responding within a given time in a unimodal auditory trial. If the probability of the response to an AV stimulus is significantly different from that predicted by the race model, integration of the auditory and visual inputs is considered to occur (Miller, 1982, 1986). The audiovisual integration of the multisensory conditions was defined by subtracting a participant's race model CDF from their AV CDF in each time bin to generate a difference curve for each participant (Miller, 1982, 1986). If their AV CDF is significantly greater than the race model CDF, the audiovisual facilitation is assumed to have occurred. Otherwise, audiovisual depression is assumed to have occurred (Meredith et al., 1987). The peak of the curve of the different CDFs is defined as the peak benefit as an important index of audiovisual integration, and the time span from the presentation of the target to the peak benefit is defined as the peak latency and is an important index of when the audiovisual integration occurred (Ren et al., 2016; Xu et al., 2020). According to previous studies (Gibney et al., 2017; Van der Stoep et al., 2015), the whole probability mass was assessed using positive area under the curve (pAUC) of differences in the CDFs. Thus, together with peak benefit, the pAUC was also calculated to assess the amount of audiovisual integration. A time interval that was significantly different in the AV CDF and the race model CDF was defined as the time window of audiovisual integration (Laurienti et al., 2006; Ren et al., 2016; Xu et al., 2020). The peak latency and the time window of audiovisual integration were used to evaluate the time point at which the audiovisual integration occurred (Laurienti et al., 2006; Ren et al., 2016; Xu et al., 2020).

RESULTS

Hit Rates

The hit rate under all conditions was greater than 85%, showing that the data were useable for further analysis (see Figure 2, Panel A). The 4 (Attentional load: no, low, medium, and high) × 3 (Target modality: A, V, and AV) ANOVA revealed that the attentional load, $F(3, 57) = 30.240$, $p < .001$, $\eta_p^2 = 0.614$, had a significant main effect, with the highest hit rate for the no and low attentional load conditions, and lowest hit rate for the high attentional load condition. However, there was no significant difference between the no and low attentional load conditions (no = low > medium > high). The stimulus type, $F(2, 38) = 6.737$, $p = .004$, $\eta_p^2 = 0.262$, also had a significant main effect, with a higher hit rate for the AV stimulus than for the A or V stimuli (all ps < .006). These results indicated that the arrangement of attentional load is reasonable, showing that with the increase of attentional load, the hit rate was significantly reduced.
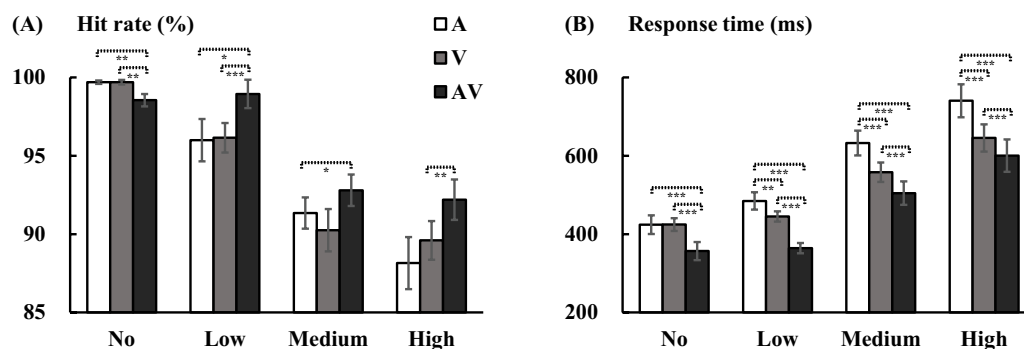
Response Time

The RTs under all conditions are displayed in Figure 2, Panel B, and the data were subjected to 4 (Attentional load: no, low, medium, and high) × 3 (Target modality: A, V, and AV) ANOVA. The results illustrated a significant main effect of attentional load, $F(3, 57) = 28.939$, $p < .0001$, $\eta_p^2 = 0.604$, showing a faster response in the no and low attentional load conditions than in the medium and high attentional load conditions (no = low > medium > high, all ps < .003). However, there was no significant difference between the no and low attentional load conditions ($p = 1.000$). There was a significant main effect of the stimulus type, $F(2, 38) = 91.017$, $p < .001$, $\eta_p^2 = 0.827$, showing that the response to the AV stimulus was significantly faster than the response to the V and A stimuli (AV > V > A, all ps < .001). Additionally, the interaction between attentional load and the stimulus type was also significant, $F(6, 114) = 10.395$, $p < .001$, $\eta_p^2 = 0.604$. The post hoc analysis for attentional load showed that, for all stimulus types, the responses were faster in the no and low attentional load conditions than in the medium and high attentional load conditions (no = low > medium > high, all p < .004). However, there was no significant difference between the no and

low attentional load conditions (all ps > .614). The post hoc analysis for attentional load showed that, in all attentional load conditions, the responses were faster for the AV stimulus than for the A and V stimuli (all ps < .001). There were no significant response differences between the A and V stimuli in the no and low attentional load conditions (all ps > .060), but there were significant differences between the A and V stimuli in the medium and high attentional load conditions (all ps < .001).

Rate Model Comparisons

To evaluate how the audiovisual integration changes with the addition of load, a race model was used to analyze the response times in each trial. Audiovisual integration was assessed using probability difference calculations by subtracting the race model CDFs from the AV CDFs under each condition. The results showed a significant audiovisual integration ($p < .05$, one-sample t test) under all four attentional load conditions (see Figure 3, Panel A for the no attentional load condition, Panel B for the low attentional load condition, Panel C for the medium attentional load condition, and Panel D for the high attentional load condition).

Overall, as shown in Table 1, the peak benefit was higher in the low attentional load condition (18.79%) than in the no (11.97%), medium (13.40%), and high attentional load conditions (6.99%), exhibiting an inverted U-shaped function. Furthermore, similar to the results of peak benefit, the pAUC was larger in the low attentional load condition (28.37 ms) than in the no (16.08 ms), medium (17.42 ms), and high attentional load conditions (12.95 ms). These results indicated that the audiovisual integration might be increased under attentional load to some degree, but it decreases inversely when the attentional load is overloaded. The peak latency for each participant was subjected to a one-way ANOVA, and the results showed a significant main effect of attentional load, $F(3, 57) = 20.235$, $p < .001$, $\eta_p^2 = 0.516$, showing that the peak latency was delayed with the addition of attentional load, showing 290 ms, 340 ms, 360 ms, and 380 ms for the no, low, medium, and high attentional load conditions, respectively. The time window of audiovisual integration was also delayed with the addition of attentional load, showing 210–370 ms, 260–490 ms, 280–450 ms, and 300–500 ms for the no, low, medium, and high attentional load

FIGURE 2.

Mean hit rates and response times with the standard errors of the mean (SEM) for each condition. The hit rate was decreased (Panel A) and the response was slow down (Panel B) with the addition of attentional load. *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.
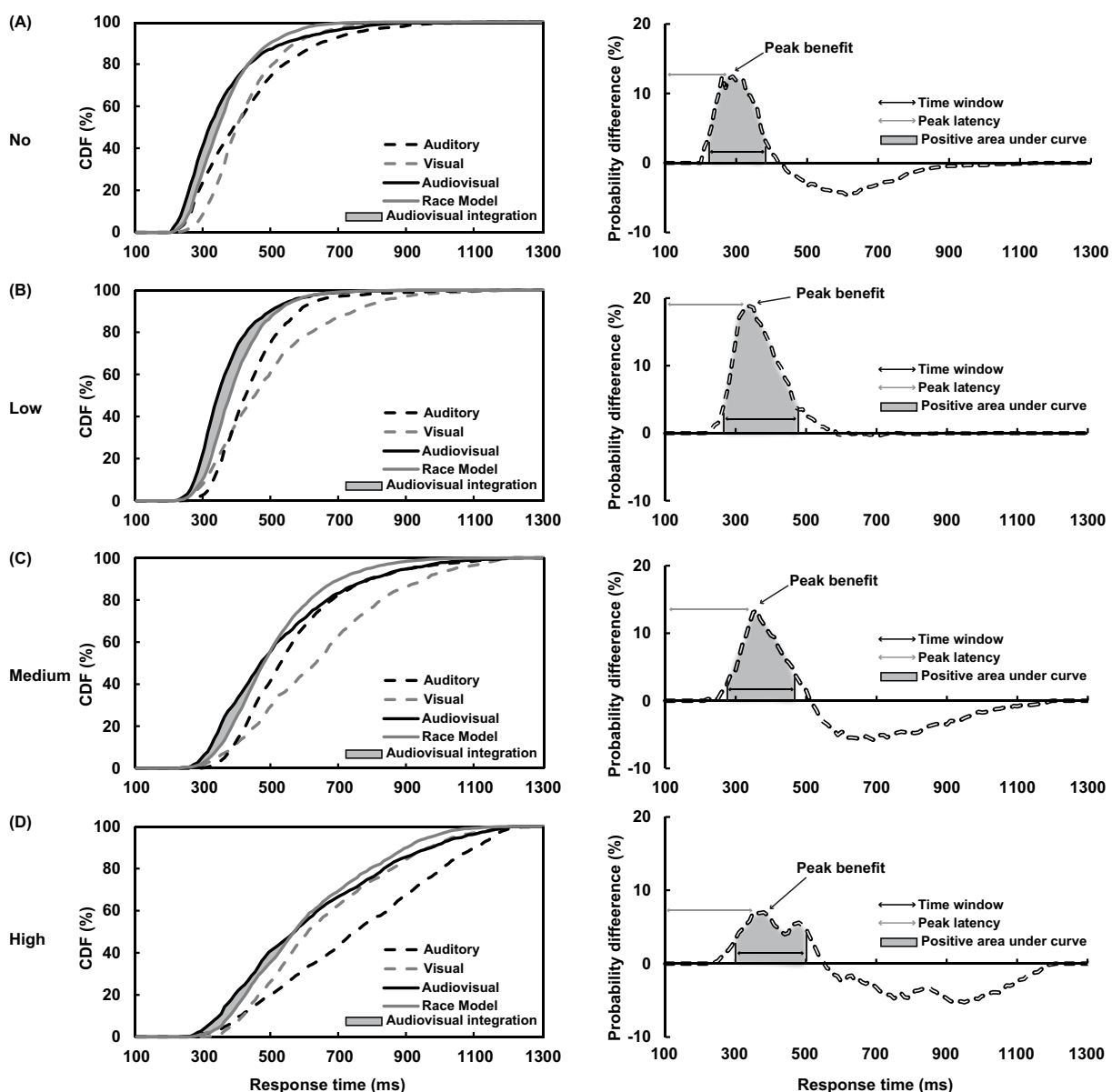
conditions, respectively. The results for the peak latency and the time window of audiovisual integration indicated that audiovisual integration was delayed with the addition of attentional load. Furthermore, a significant audiovisual depression effect was found for the no (500–520 ms), medium (610–1020 ms), and high (710–1160 ms) attentional load conditions, but not for the low attentional load condition.

DISCUSSION

The aim of the present study was to investigate the effect of auditory attentional load on audiovisual integration. The audiovisual integration exhibited an inverted U-shaped function, and audiovisual integration was delayed with increasing attentional load. Additionally, an audio-

visual depression effect was found in the no, medium, and high attentional load conditions, but not in the low attentional load condition.

These findings are consistent with our hypothesis that audiovisual integration increases with the addition of auditory attentional load in the low attentional load condition but decreases inversely owing to the competition for attentional resources during the RSAP task in the high auditory attentional load condition, exhibiting an inverted U-shaped function. Attentional resources include three separate networks: alerting, orienting, and executive control, with alerting being defined as achieving and maintaining an alert state (Petersen & Posner, 2012; Posner & Petersen, 1990). In the low auditory attentional load condition, although the participants were instructed to respond only to the auditory/visual discrimination task, more alerting resources were

FIGURE 3.

Significant audiovisual integration effects were found under the no attentional load (Panel A); low attentional load (Panel B); medium attentional load (Panel C); and high attentional load (Panel D) conditions.

TABLE 1.

Peak Benefit, pAUC, Peak Latency, and Time Window of Audio-visual Integration under Each Attentional Load Condition

	No load	Low load	Medium load	High load
Peak benefit (%)	11.97	18.79	13.40	6.99
pAUC (ms)	16.08	28.37	17.42	12.95
Peak latency (ms)	290	340	360	380
Time window (ms)	210–390	260–490	290–480	300–500
	500–520	—	610–1020	710–1160

Note. pAUC: positive area under curve of differences of CDFs.

activated than in the no attentional load condition while simultaneously detecting the stimuli of the auditory/visual discrimination task and suppressing the distractors of the RSAP task (Ho et al., 2009; Ren, Zhou, et al., 2020). According to the arousal hypothesis (proposed by Yerkes & Dodson, 1908), moderately difficult tasks activate the best arousal level, as characterized by an inverted U-shaped function. With increasing physical arousal, enhanced performance is predicted to an optimal point and then decreases with further arousal (Kamijo et al., 2007; Yerkes & Dodson, 1908). Therefore, in the low auditory attentional load condition, the amount of activated attentional resources is larger than that of the no attentional load condition. Studies have confirmed that audiovisual integration is higher for attended locations than for unattended locations (Talsma et al., 2010; Talsma et al., 2009; Talsma & Woldorff, 2005). Therefore, the higher audiovisual integration in the low auditory attentional load condition than in the no attentional load condition is attributed to the greater activation of attentional resources. In the medium auditory attentional load condition, the participants were instructed to respond to the auditory/visual discrimination task and the RSAP task equally. The response to the RSAP task requires extra attentional resources (Ho et al., 2009; Ren, Zhou, et al., 2020). Therefore, the RSAP distractor demanded more attentional resources in the medium auditory attentional load condition than in the low auditory attentional load condition. Fewer attentional resources were left to process the stimuli of the auditory/visual discrimination task in the medium auditory attentional load condition than in the low auditory attentional load condition, which led to lower audiovisual integration in the medium load condition than in the low load condition. Similarly, as the complexity of the RSAP task increases, more attentional resources are allocated to the RSAP task, and fewer attentional resource are left to process the auditory/visual discrimination task, as shown by the decreased hit rate in the high attentional load condition. Therefore, audiovisual integration was lower in the high auditory attentional load condition than in the medium auditory attentional load condition.

These results were also consistent with those of Alsius et al. (2005; Alsius et al., 2014) and Ren, Li et al. (2020). In these experiments, McGurk words and the distractor stimuli were presented simultaneously. Under the low attentional load condition, the participant was instructed only to repeat the McGurk words while ignoring the simultaneous distractor stimuli (single task). Under the high attentional load condition, the participants were asked to simultaneously repeat

the McGurk words and respond to the target of the distractor stimuli by pressing the mouse button (dual task). By contrast, in Ren, Li et al. (2020)'s study, only the auditory/visual discrimination task was presented in the no attentional load condition, which was used to evaluate audiovisual integration (single task). In the low visual attentional load condition of that study, both the auditory/visual discrimination task and the RSVP task were presented simultaneously, but the participants were instructed only to respond to the auditory/visual discrimination task (single task). In the high attentional load condition, the auditory/visual discrimination task and the RSVP task were presented simultaneously, and the participants were instructed to respond to both tasks (dual task). However, the results of Alsius et al. (2005; Alsius et al., 2014) and Ren, Li et al. (2020) were consistent in that the audiovisual integration was higher in the low attentional load condition than in the high attentional load condition. However, only Ren, Li et al. found higher audiovisual integration in the low visual attentional load condition than in the no attentional load condition. Considering these findings and the results of the present study on auditory attention, we propose that there may be a positive correlation between audiovisual integration and the attentional resources occupied by in the auditory/visual discrimination task when there is attentional competition from distractors. However, further neuroimaging studies are needed. Notably, in the no and low attentional load conditions, a single task was conducted, but there were dual tasks in the medium and high attentional load conditions. Furthermore, the amount of stimuli was different between the no and low attentional load condition, and the secondary cognitive task was different between the medium and high attentional load conditions. Therefore, task-dependent effects might also have contributed to the changes in audiovisual integration in the various attentional load conditions, which should be given full consideration in the future studies.

The audiovisual integration was delayed with the addition of attentional load. Previous studies have shown that the response speed to target stimuli is reduced by irrelevant distractors (Fan et al., 2009; Fan et al., 2002). Therefore, the response speed in the low auditory attentional load condition was slower than that in the no attentional load condition. In the present study, with increased attentional load, the cognitive processing depth and the task difficulty were also increased. In the medium auditory attentional load condition, the participants were instructed to make a category discrimination for the RSAP stream. However, an accurate character identification was required in the high auditory attentional load condition, and this condition required more sophisticated processing. The more complex and difficult the information processing, the more time is needed to make a response (Goldhammer et al., 2014; Houlihan et al., 1998). Therefore, the response speed slowed with increased attentional load. Colonius and Diedrich (2004; Diedrich et al., 2008) proposed a "time window of integration model," and they presumed that cross-modal information integration includes at least two serial stages of saccadic reaction times: an early afferent stage of peripheral processing (first stage) and a compound stage of converging subprocesses (second stage). The first stage consists of very early sensory processing, and the processing time

is assumed to be independent for unimodal sensory stimuli. If the information processing in the first stage terminates within a given time interval, the integration of auditory and visual information is assumed to occur in the second stage. In the current study, as the attentional load increased, the RT was significantly reduced, leading to a delayed termination of the first stage and then delayed audiovisual integration in the second stage (Ren, Guo, et al., 2020; Ren et al., 2016).

The audiovisual depression effect occurred in the no, low, and high attentional load conditions, but not in the low attentional load condition. As in the arousal hypothesis, the optimal attentional load led to maximum physical arousal (Yerkes & Dodson, 1908). In the low attentional load condition, neural activation and physical arousal arrived in a relatively optimal activation state, exhibiting hyperexcitability to various external stimuli. This hyperexcitability then contributed to the super-additivity of auditory and visual stimuli and activated the facilitation system (Driver & Noesselt, 2008; Meredith et al., 1987; Stein & Stanford, 2008). However, in the no attentional load condition, the simple task induced slack and inattention, which led to suppression as well as facilitation (Kamijo et al., 2007; Yerkes & Dodson, 1908). In the medium and high attentional load conditions, the participants were instructed to respond to the auditory/visual discrimination task and the RSAP task simultaneously. While integrating the auditory and visual stimuli, the individuals must allocate attentional resources to suppress the merging of irrelevant RSAP stimuli (Petersen & Posner, 2012; Posner & Petersen, 1990), and both the facilitation and suppression systems were activated. Therefore, both facilitation and suppression effects were found in the no, low, and high attentional load conditions.

It is also notable that Wahn et al. (2017) reported interference was significant when the secondary task was performed simultaneously, showing a higher detection threshold for dual tasks than for a single task. Therefore, in the current study, the simultaneous secondary RSAP task influenced the perception of the auditory and visual stimuli of the auditory/visual discrimination task resulting from the reduction of the signal-to-noise ratio. This influence should be optimized in the future studies.

REFERENCES

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

Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., & Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Frontiers in Psychology, 5*. doi: 10.3389/fpsyg.2014.00727

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*, 839–843. doi: 10.1016/j.cub.2005.03.046

Arrighi, R., Lunardi, R., & Burr, D. (2011). Vision and audition do not share attentional resources in sustained tasks. *Frontiers in Psychology, 2*, 56. doi: 10.3389/fpsyg.2011.00056

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience, 16*, 1000–1009. doi: 10.1162/0898929041502733

Diederich, A., Colonius, H., & Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia, 46*, 2556–2562. doi: 10.1016/j.neuropsychologia.2008.03.026

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron, 57*, 11–23. doi: 10.1016/j.neuron.2007.12.013

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

Driver, J., & Spence, C. (1998b). Crossmodal attention. *Current Opinion in Neurobiology, 8*, 245–253. doi: 10.1016/S0959-4388(98)80147-5

Eimer, M. (1996). ERP modulations indicate the selective processing of visual stimuli as a result of transient and sustained spatial attention. *Psychophysiology, 33*, 13–21. doi: 10.1111/j.1469-8986.1996.tb02104.x

Eimer, M., & Forster, B. (2003). Modulations of early somatosensory ERP components by transient and sustained spatial attention. *Experimental Brain Research, 151*, 24–31. doi: 10.1007/s00221-003-1437-1

Fan, J., Gu, X., Guise, K. G., Liu, X., Fossella, J., Wang, H., & Posner, M. I. (2009). Testing the behavioral interaction and integration of attentional networks. *Brain and Cognition, 70*, 209–220. doi: 10.1016/j.bandc.2009.02.002

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*, 340–347. doi:

10.1162/089892902317361886

Gibney, K. D., Aligbe, E., Eggleston, B. A., Nunes, S. R., Kerkhoff, W. G., Dean, C. L., & Kwakye, L. D. (2017). Visual distractors disrupt audiovisual integration regardless of stimulus complexity. *Frontiers in Integrative Neuroscience, 11*. doi: 10.3389/fnint.2017.00001

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*, 608–626. doi: 10.1037/a0034716

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334–337. doi: 10.1038/383334a0

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: When spatial correspondence matters. *Experimental Brain Research, 195*, 261–272. doi: 10.1007/s00221-009-1778-5

Houlihan, M. E., Stelmack, R. M., & Campbell, K. B. (1998). Intelligence and the effects of perceptual processing demands, task difficulty and processing speed on P300, reaction time and movement time. *Intelligence, 26*, 9–25. doi: 10.1016/S0160-2896(99)80049-X

Kamijo, K., Nishihira, Y., Higashiura, T., & Kuroiwa, K. (2007). The interactive effect of exercise intensity and task difficulty on human cognitive processing. *International Journal of Psychophysiology, 65*, 114–121. doi: 10.1016/j.ijpsycho.2007.04.001

Laura, B., Kenneth, C. R., Roy, E. C., Daniel, H. W., Woldorff, M. G., & Purves, D. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 18751–18756. doi: 10.1073/pnas.0507704102

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27*, 1155–1163. doi: 10.1016/j.neurobiolaging.2005.05.024

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 451–468. doi: 10.1037/0096-1523.21.3.451

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics, 56*, 183–197. doi: 10.3758/BF03213897

Mcdonald, J. J., Tedersälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*(6806), 906–908. doi: 10.1038/35038085

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. doi: 10.1038/264746a0

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience, 7*(3215–3229. doi: 10.1097/00005072-199808000-00008

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247–279. doi: 10.1016/0010-0285(82)90010-X

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40*, 331–343. doi: 10.3758/

BF03203025

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*, 154–163. doi: 10.1016/S0926-6410(03)00089-2

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Reviev of Neuroscience, 35*, 73–89. doi: 10.1146/annurev-neuro-062111-150525

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Reviev of Neuroscience, 13*, 25–42. doi: 10.1146/annurev.ne.13.030190.000325

Ren, Y., Guo, A., Xu, Z., Wang, T., Wu, R., & Yang, W. (2020). Age-related functional brain connectivity during audio–visual hand-held tool recognition. *Brain and Behavior, n/a(n/a), e01759*. doi: 10.1002/brb3.1759

Ren, Y., Li, S., Wang, T., & Yang, W. (2020). Age-related shifts in theta oscillatory activity during audio-visual integration regardless of visual attentional load. *Frontiers in Aging Neuroscience, 12*: 571950. doi: 10.3389/fnagi.2020.571950

Ren, Y., Ren, Y., Yang, W., Tang, X., Wu, F., Wu, Q., Takahashi, S., Ejima, Y., & Wu, J. (2018). Comparison for younger and older adults: Stimulus temporal asynchrony modulates audiovisual integration. *International Journal of Psychophysiology, 124*, 1–11. doi: 10.1016/j.ijpsycho.2017.12.004

Ren, Y., Yang, W., Nakahashi, K., Takahashi, S., & Wu, J. (2016). Audiovisual integration delayed by stimulus onset asynchrony between auditory and visual stimuli in older adults. *Perception, 46*, 205. doi: 10.1177/0301006616673850

Ren, Y., Zhou, Z., Bi, J., Li, J., Wang, T., & Yang, W. (2020). Effect of attentional load on audio-visual integration:an ERP study [Conference presentation]. International Conference on Mechatronics and Automation, Beijing. doi: 10.1109/ICMA49215.2020.9233743

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 6614*(385). doi: 10.1038/385308a0

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*, 147–152. doi: 10.1016/S0926-6410(02)00069-1

Spence, C. (2010a). Crossmodal attention. *Scholarpedia, 5*, 6309. doi: 10.4249/scholarpedia.6309

Spence, C. (2010b). Crossmodal spatial attention. *Annals of the New York Academy of Sciences, 1191*, 182–200. doi: 10.1111/j.1749-6632.2010.05440.x

Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*. Oxford University Press.

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology, 13*, R519–521. doi: 10.1016/s0960-9822(03)00445-7

Stein, B. E. (2012). *The new handbook of multisensory processing*. MIT Press.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8*, 497–506. doi: 10.1162/jocn.1996.8.6.497

Stein, B. E. & Meredith, M. A. (1993). *The merging of the sense*. The MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*, 255–266. doi: 10.1038/nrn2331

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex, 17*, 679–690. doi: 10.1093/cercor/bhk016

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*, 400–410. doi: 10.1016/j.tics.2010.06.008

Talsma, D., Senkowski, D., & Woldorff, M. G. (2009). Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. *Experimental Brain Research, 198*, 313–328. doi: 10.1007/s00221-009-1858-6

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience, 17*, 1098–1114. doi: 10.1162/0898929054475172

Van der Stoep, N., Van der Stigchel, S., & Nijboer, T. C. W. (2015). Exogenous spatial attention decreases audiovisual integration. *Attention, Perception, & Psychophysics, 77*, 464–482. doi: 10.3758/s13414-014-0785-1

Wahn, B., & König, P. (2015). Audition and vision share spatial attentional resources, yet attentional load does not disrupt audiovisual integration. *Frontiers in Psychology, 6*: 1084. doi: 10.3389/fpsyg.2015.01084

Wahn, B., & König, P. (2017). Is attentional resource allocation across sensory modalities task-dependent? *Advances in Cognitive Psychology, 13*, 83–96. doi: 10.5709/acp-0209-2

Wahn, B., Murali, S., Sinnett, S., & König, P. (2017). Auditory stimulus detection partially depends on visuospatial attentional resources. *i-Perception, 8*, 2041669516688026. doi: 10.1177/2041669516688026

Xu, Z., Yang, W., Zhou, Z., & Ren, Y. (2020). Cue–target onset asynchrony modulates interaction between exogenous attention and audiovisual integration. *Cognitive Processing.* doi: 10.1007/s10339-020-00950-2

Yang, W., Yang, J., Gao, Y., Tang, X., Ren, Y., Takahashi, S., & Wu, J. (2015). Effects of sound frequency on audiovisual integration: an event-related potential study. *Plos One, 10*, e0138296. doi: 10.1371/journal.pone.0138296

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482. doi: 10.1002/cne.920180503