Title of Thesis

Psychophysical Studies on Modulation Mechanism of Attentional Load in Audiovisual Integration

March 2021

Qingqing Li

The Graduate School of

Natural Science and Technology

(Doctor's Course)

OKAYAMA UNIVERSITY

Abstract

We live in a world, which is rich in information from many sensory modalities (e.g., auditory, visual, olfactory, and somatosensory). The brain can screen available information from multiple senses and integrate them to better perceive the external environment, thereby shaping and guiding our behaviours. Most of the information that humans use to recognize the world is derived from auditory and visual modalities. For example, when one tries to localize a singing bird flitting between the branches of a tree with luxuriant foliage, the combination of auditory and visual input information - as compared to using only auditory or visual information - will probably increase the accuracy and speed of the localization process. The phenomenon by which stimuli from visual and auditory sensory organs can be integrated into a coherent representation to better perceive information is called audiovisual integration (AVI). Although it is generally believed that attention plays a complex and multifaceted role in the integration of input from different sensory modalities, whether AVI is affected by top-down attentional load remains less clear. Additionally, stimulus congruency (e.g., semantic congruency, spatial congruency) may be proposed as a factor that determines the extent of attentional effects on AVI. Therefore, the main aim of this present thesis was to investigate how attentional load interact with stimulus congruency to influence AVI.

In part 1, we rigorously examine how semantic congruency interacts with attentional load to influence the AVI of common objects by applying a dual-task paradigm. Currently, many studies are beginning to use a dual-task paradigm in which a distracter task is adopted to modulate the levels of the endogenous attentional resources available for the secondary task to explore the effect of attentional load on AVI processing. In present study, we adopted a rapid serial visual presentation (RSVP) task as the distractor task to impose different levels of attentional load, namely, no load, low load, and high load. Specifically, participants were instructed to ignore the presented RSVP stream under no load condition, while participants simultaneously performed the AVI task and a distractor task that required them to search a central RSVP stream for either a yellow letter (low load) or a white digit (high load). The AVI was assessed by adopting an animal identification task using unisensory (animal images and sounds) and AV stimuli (semantically congruent AV stimuli and semantically incongruent AV stimuli). The results confirmed that attentional load did not attenuate the integration of semantically congruent AV stimuli. However, semantically incongruent animal sounds and images were not integrated (as there was no multisensory facilitation), and the interference effect produced by the semantically incongruent animal sounds and images was reduced by increased attentional load manipulations. We further observed an asymmetric cross-modal interference effect supporting the visual dominance hypothesis; specifically, the auditory distractor effect was stronger than the visual distractor effect under all attentional load conditions. These findings highlight the critical role of semantic congruency in modulating the effect of attentional load on the AVI of common objects.

In part 2, to further clarify how the cross-modal interaction of AV stimuli is influenced by increased attentional load when attention is only focused on visual modality (selective attention to visual modality), and whether semantic association

between AV stimuli modulates the effect of increased attentional load on the AVI of common objects. We manipulated the amount of available attentional resources by applying a dual-task paradigm and constructed three attentional load levels (no load, low load, and high load) by using a rapid serial visual presentation (RSVP) task. Additionally, individuals are instructed to ignore the auditory stimuli, and only response to visual target in the AVI task. And semantic associations between AV stimuli were composed of animal pictures presented concurrently with either semantically congruent, incongruent or unrelated auditory stimuli. The results showed that attentional load did not reliably alter the amount of the auditory enhancement effects caused by semantically congruent AV stimuli on this task. However, attentional load disrupts the auditory enhancement effects of the semantically unrelated and incongruent AV stimuli. These findings suggested that the strong semantic associations between AV stimuli played an important role in withstanding the effect of attentional load on AVI of modality-specific selective attention.

In part 3, we explored whether increased attentional loads would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions. We will adopt an RSVP stream as the distractor task to manipulate different attentional load: no load, low load, and high load. Specifically, participants simultaneously performed the AVI task and a distractor task that required them to search a central RSVP stream for either a red letter (low load) or two different coloured letters (high load) in a series of different coloured characters. Under no load condition, participants were only asked to response to AVI task. In the AVI task, participants were instructed to respond to a specific image

(black–white checkboard with two black dots) while ignoring all sounds (i.e., pure tone and white noise). And spatial congruency was controlled by presenting visual and auditory stimuli in the same or different locations. The results showed that significant integration of spatial congruent AV stimuli occurred regardless of attentional load; however, increased attentional load reduced the integration of spatial incongruent AV stimuli. These findings highlight the critical role of spatial congruency in modulating the effect of attentional load on the integration of simple and arbitrarily AV stimuli.

In conclusion, we found that whether AVI is influenced by increased attentional load conditions depends on stimulus congruency between AV stimuli. Specifically, when using complex naturalistic common objects which corresponds to semantic content and operates on a higher level, our results show that semantic congruency plays a critical role in modulating the effect of attentional load on this AVI processing, whether or not in distributed attention or focused visual attention settings. Further, when not using complex naturalistic common objects, but presenting simple and arbitrarily paired bimodal stimuli, we found that attentional load did not attenuate the integration of spatially congruent AV stimuli, but disrupted the integration of spatially incongruent AV stimuli. Thus, it seems that stimulus congruency plays a critical factor in modulating the effect of attentional load on the AVI.

Key words: Audiovisual integration, Attentional load, Semantic congruency, Spatial congruency, Selective attention.

Table of Contents

Chapter 1 Introduction

Summary

This chapter introduces the concept of audiovisual integration, stimuli congruency and attentional load. The previous studies of how attentional load influences audiovisual integration has also been summarized here. Additionally, the data analysis technique of Race Modal has been introduced. At last, the purpose and contents of the thesis are briefly described.

1.1 Audiovisual Integration

Multisensory information combined from different sensory channels performs faster and more accurate in discrimination and identification tasks than unimodal stimuli, and thus results in better response advantages in the neural, behavioral and perceptual realms [1-2]. Moreover, the brain can perform the integration processing tasks across different sensory channels, such as the audio-visual (AV) domain [3], and it can merge the AV inputs into an overall more coherent perception.

Audiovisual integration has been demonstrated to occur in several different brain areas at different stages of sensory processing using different stimulus types [4]. For example, the audiovisual integration of simple and arbitrarily paired bimodal stimuli occurs at relatively low cortical areas, whereas the audiovisual integration of complex stimuli, especially high-level semantic stimuli, likely occurs at higher cortical areas [4-5]. The typical associations between complex auditory and visual stimuli at the level of semantic content rely on the audiovisual (AV) integration of common objects [5]. The AV integration of common objects, which involves interactions between a complex visual stimulus and a sound counterpart of living and non-living familiar objects, such as the binding of the picture of a dog and a corresponding barking sound, closely corresponds to semantic content and operates on a higher level [5-6]. Furthermore, it has been proposed that stimulus congruency and attention are two important factors that can influence cross-modal integration [7].

1.1.1 Stimulus congruency

Stimulus congruency represents the correlation or accordance between stimulus characteristics [5], such as spatiotemporal concordance [8] and semantic congruency [9] , etc.

Spatial congruence was the locational correspondence of incoming signals from different sensory channels [10], and it can furtherly contribute multisensory information to produce facilitation effect [7]. Spence et al (2004) have found that spatial congruency has an impact on cross-modal visual-tactile integration processing (ignore visual) [11]. Furthermore, attention may have different effects on the processing of audiovisual stimuli presented at the same (an early influence of attention) or different (later attentional modulations) spatial positions [11-12]. It has been indicated that the spatial congruence of bimodal audio-tactile cue plays an important role in helping resist the interference of attentional load [13]. Specifically, when bimodal stimuli are presented in the same location, they effectively attract spatial attention even under high attentional load [13-14], if the location of tactile stimuli is neutral to auditory stimuli, the bimodal stimuli will not produce a spatial cuing effect or capture attention regardless of attentional load [13]. However, it is not at all clear whether attentional load would have different effects on the cross-modal interaction of audiovisual stimuli presented at the same or different spatial positions.

"Semantic congruency is the learned associations between the individual sensory elements of a single object or event" [15], such as the images and sounds belonging to common objects[5] and the letters and pronunciations of the same alphabet[9]. Moreover, many studies have found that, in addition to simple features such as spatial and temporal correspondence [11-12], the semantic congruency of multisensory stimuli also plays a key role in determining how the nervous system handles the stimuli [15]. Furthermore, while it has been suggested that semantically congruent bimodal stimuli may produce better behavioral performances than unimodal stimuli, no enhanced effect is found for semantically incongruent AV stimuli [8,15]. Nonetheless, because the processing of semantic relevance between multiple sensory stimuli involves higher level cognitive processing and has deep interrelationships with attention [8,9,15], memory [16-17], etc. So, it is necessary to explore how semantic congruency interacts with other high-level processing mechanisms to influence the cross-modal integration processing.

1.1.2 Attentional loads

In fact, the role of attention in the integration of input from different sensory modalities is complex and multifaceted [18-19], and whether the occurrence of multisensory integration is relatively automatic and not affected by top-down attentional control has become an ongoing debate [20-21]. Currently, many studies are beginning to use a dual-task paradigm in which a distracter task is adopted to modulate the levels of the endogenous attentional resources available for the secondary task to explore the effects of attentional load on multisensory integration processing.

According to attentional load theory, although attentional resources are limited in capacity, they can be used to process all received stimuli until all available resources are exhausted [22-23]. Moreover, it has been shown that dual task designs reduce the attentional capacity of main task [24] because dividing attention between two concurrent tasks results in a decrement in behavioral performance relative to when only the main task is performed [26]. Accordingly, dual task designs can be adopted by constructing different attentional loads to explore whether the absence of attentional resources influences the audiovisual integration.

Although the effect of attentional load on the cognitive performance of stimulus congruency has been investigated in the cross-modal integration field, the behavioral evidence of cross-modal integration related enhancement effect is mixed. For example, while some findings indicate that when bimodal stimuli are presented in the same location, they effectively attract spatial attention even under high attentional load [13-14], if the location of tactile stimuli is neutral to auditory stimuli, the bimodal stimuli will not produce a spatial cuing effect or capture attention regardless of attentional load [13]. In a similar vein, it has been demonstrated that a static face with happy emotional states receives more happy responses on auditory emotion judgments than a fearful face even when paired with different attentional load tasks [27]. In addition, it has been demonstrated that while attentional load interferes with the cross-modal speech integration of incongruent stimuli (McGurk effect) [28-30], it will not interfere with the congruent related enhancement effect produced by speech AV stimuli [30]. These results maybe suggest that the cross-modal integration related enhancement effect does not always occur in the absence of attentional resources and that stimulus congruency may modulate the effect of attentional load on the cross-modal integration processing. However, it remains unclear how stimulus congruency interacts with attentional load to influence the audiovisual integration.

1.2 The purpose of the present thesis

The main aim of this present thesis was to investigate how attentional load interacts with stimuli congruency to influence the audiovisual integration.

Chapter 1 introduces the concept of audiovisual integration, semantic congruency and attentional load. The previous studies of how stimuli congruency and attentional load influence the audiovisual integration have also been summarized here. At last, the purpose and contents of the thesis are briefly described.

Chapter 2 describes how semantic congruency interacts with attentional load to influence the AV integration of common objects by applying a dual-task paradigm to rigorously.

Chapter 3 describes whether the cross-modal interaction of AV stimuli can occur automatically, and is not restricted by increased attentional loads when attended selectively to visual modality, and whether semantic association among AV stimuli modulates the effect of increased attentional loads on the AV integration of common objects in the setting of attended selectively to visual modality.

Chapter 4 describes whether the cross-modal interaction of AV stimuli can occur automatically, and is not restricted by limited attentional resource in the setting of focused visual attention, and whether attentional load would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions.

Chapter 5 present general conclusions based on the findings of the three experiments. And the future challenges are also described.

Chapter 2 Semantic Congruency Modulates the Effect of Attentional Load on Audiovisual Integration of Common Object

Summary

Attentional processes play a complex and multifaceted role in the integration of input from different sensory modalities. However, whether increased attentional load disrupts the audiovisual (AV) integration of common objects that involve semantic content remains unclear. Furthermore, knowledge regarding how semantic congruency interacts with attentional load to influence the AV integration of common objects is limited. We investigated these questions by examining AV integration under various attentional load conditions. Audiovisual integration was assessed by adopting an animal identification task using unisensory (animal images and sounds) and AV stimuli (semantically congruent AV objects and semantically incongruent AV objects), while attentional load was manipulated by using an RSVP task. Our results indicate that attentional load did not attenuate the integration of semantically congruent AV objects. However, semantically incongruent animal sounds and images were not integrated, and the interference effect produced by the semantically incongruent AV objects was reduced by increased attentional load manipulations. These findings highlight the critical role of semantic congruency in modulating the effect of attentional load on the AV integration of common objects.

2.1 Background

In daily life, individuals usually receive information from many sensory modalities, and the human brain can combine and bind the available information from multiple senses to better perceive the external environment. The phenomenon by which stimuli from multiple sensory organs can be integrated into a coherent representation to better perceive information is called multisensory integration [1-2]. Multisensory integration has been demonstrated to occur in several different brain areas at different stages of sensory processing using different stimulus types [4]. For example, the multisensory integration of complex stimuli, especially high-level semantic stimuli, likely occurs at higher cortical areas [4-5]. Moreover, the typical associations between complex auditory and visual stimuli at the level of semantic content rely on the audiovisual (AV) integration of common objects [5,31].

The AV integration of common objects, which involves interactions between a complex visual stimulus and a sound counterpart of living and non-living familiar objects, such as the binding of the picture of a dog and a corresponding barking sound, closely corresponds to semantic content and operates on a higher level [5]. Moreover, recent studies have shown that semantic congruency, which modulates the semantic association between the individual sensory elements of a single object [32], has an impact on AV integration. Specifically, bimodal stimuli conveying semantically congruent information can be preferentially selected to improve behavioural performance, whereas incongruent bimodal stimuli impair performance [33-35]. At the neural level, it has been reported that the integration of semantically congruent AV combinations of common objects evokes stronger activations of posterior temporal regions around the STS (pSTS) and middle temporal gyrus (MTG) than incongruent combinations [3]. Furthermore, the processing of semantic congruency between the unimodal components of a multisensory signal involves higher level cognitive processing, and semantic congruency was proposed as a factor that determines the extent of attentional effects on AV integration [9,34,36]. Specifically, using spoken and written nouns in a target detection task, Mishra and Gazzaley (2012) showed that compared to selective attention to either the visual or the auditory modality, distributing attention across both auditory and visual domains enhances performance for congruent AV stimuli, but resolves interference for incongruent AV stimuli [36]. Thus, it seems that the integration of semantically congruent AV stimuli may be less susceptible to top-down attentional controls than the interference effect of incongruent AV stimuli.

In fact, the role of attention in the integration of input from different sensory modalities is complex and multifaceted [18-19], and whether the occurrence of multisensory integration is relatively automatic and not affected by top-down attentional control has become an ongoing debate [20-21]. Currently, many studies are beginning to use a dual-task paradigm in which a distracter task is adopted to modulate the levels of the endogenous attentional resources available for the secondary task to explore the effects of attentional load on multisensory integration processing. Using this approach, it has been demonstrated that the “ventriloquist effect” (the temporal integration of simple audiovisual stimuli) is not influenced by attentional load, Specifically, a shift in auditory localization toward peripheral flashes can still be found regardless of whether attention was exogenously directed away from the flashes [27]. In a similar manner, some findings indicate that multisensory cues can more effectively attract spatial attention even under high attentional load than unimodal cues, indicating that the spatial integration of simple multisensory cues is not affected by increased attentional demands [13-14]. In contrast, some results have demonstrated that attentional load severely interfered with AV speech integration as indexed by the McGurk effect, in which a speech sound paired with an incongruent lip movement leads to a fused speech sound [28-30]; this type of speech perception is usually considered highly complex and requires extensive neural processing [30]. Nevertheless, although these studies have obtained contradictory experimental findings, they investigated different aspects of multisensory integration (temporal or spatial integration of simple multisensory stimuli; AV speech perception). Furthermore, it seems that several aspects related to the impact of attentional load on multisensory integration have not been fully studied; specifically, it remains an open question whether attentional load also disrupts AV integration of common objects. Moreover, how semantic congruency interacts with attentional load to influence the AV integration of common objects also remains unclear.

Thus, the purpose of the current study was to apply a dual-task paradigm to rigorously examine how semantic congruency interacts with attentional load to influence the AV integration of common objects. The dual-task paradigm reduces the attentional capacity dedicated to the main task because dividing attention between two concurrent tasks results in a decrease in behavioural performance relative to when only the main task is performed [26]. In addition, a distractor task of low difficulty allows the allocation of spare attentional resources to another simultaneous task; however, performing a highly difficult distractor task may exhaust the attentional resources that can be allocated to another task [22-23]. Thus, by increasing the difficulty of the distractor task, attentional load can be controlled at different levels. We adopted a rapid serial visual presentation (RSVP) task as the distractor task to impose different levels of attentional load, namely, no load, low load, and high load. In addition, we also controlled the semantic congruency in the AV integration task by adopting semantically congruent AV objects (e.g., dogs with barks) and semantically incongruent AV objects (e.g., birds with barks) of common objects. Finally, our hypotheses were as follows: (1) the integration of semantically congruent AV object features would not be significantly attenuated by increased attentional load; (2) however, the multisensory interference effect of semantically incongruent AV object features would be significantly decreased by increased attentional load. Our behavioural results are evaluated from the perspective of these hypotheses.

2.1 Methods

2.2.1 Subjects

A total of 20 volunteers (five females, mean age of 25 years) participated in this study. The participants reported normal or corrected-to-normal hearing and vision. All participants provided written informed consent, and the study procedures were approved in advance by the ethics committee of Okayama University. Two participants were excluded from further analyses due to poor data quality, specifically because they had low average accuracy of the audiovisual integration task even under the no-load condition (70% accuracy). Therefore, data from eighteen subjects were analysed (4 females; mean age 26 years, ranging from 18 to 31 years).

2.2.2 Stimuli

All study procedures were completed in a dimly lit, electrically shielded and sound-attenuated room, specifically, a laboratory room at Okayama University, Japan. Each participant positioned his or her head on a chin rest. All visual stimuli were presented on a 24-inch VG 248LCD monitor (made by ASUS, Taiwan) with a screen resolution of 1920×1080 and a refresh rate of 144 Hz set at a viewing distance of 57 cm from the participant. Auditory stimuli were presented through speakers located on the central monitor. Additionally, two speakers (Harman/Kardon HK206, frequency response: 90-20,000 Hz) were used to present the auditory stimuli. MATLAB software (R2014b, MathWorks, MA, Psychtoolbox-3) was used to present the experimental stimuli and record the participants' responses.

We administered the animal identification task (AV integration task) with the following four basic stimulus types, each presented with equal probability: (i) sounds alone, (ii) pictures alone, (iii) paired pictures and sounds belonging to the same animal, and (iv) paired pictures and sounds belonging to different animals. The images included line drawings of a dog, bee, frog, bird, and pig developed by the Snodgrass and Vanderwart set [37] and were standardized by familiarity and complexity. All visual stimuli were presented on the lower left or lower right quadrant of the screen for 300 ms (subtending a 12° visual angle to the left or right of the centre and a 5° angle below the central fixation point).

The sounds of these five animals were collected through internet searches (http://sc.chinaz.com/tag_yinxiao/DongWuJiaoSheng.html.) and later standardized and modified such that each single animal sound had a duration of 300 ms. The animal sounds were presented at a comfortable listening level of the ~75 dB sound pressure level (SPL). Furthermore, in addition to the unimodal stimuli (animal pictures alone and animal sounds alone), the pictures and sounds of animals were also combined to form both congruent pairs (combinations of pictures and sounds belonging to the same animal) and incongruent pairs (combinations of pictures and sounds belonging to different animals). Of note, the images or sounds of a "bird" served as the target stimuli. Participants were asked to react as fast as possible to a target object ("bird") presented in the visual and/or auditory modality and to inhibit a distractor object (go/no go task). It was further explained to them that they also had to respond to semantically incongruent stimuli, in which only the visual or the auditory element was the target. Finally, five target stimulus types and four nontarget stimulus types were derived from the four basic stimulus types (Fig. 1). The five target stimulus types were as follows: visual target (V+, a picture of a bird), auditory target (A+, the tweet of a bird), a picture and sound pair in which both were targets (V+A+, a picture of a bird and the tweet of a bird), a picture and sound pair in which only the picture was a target (A-V+; e.g., a picture of a bird and the bark of a dog), and a picture and sound pair in which only the sound was a target (A+V-; e.g., a picture of a dog and the tweet of a bird).

The four nontarget stimulus types were as follows: an animal picture (V-), an animal sound (A-), a paired picture and sound of the same animal (congruent A-V-), and a picture of one animal paired with the sound of another animal (incongruent V-A-). Thus, there were nine total trial types (five target stimulus types: V+, A+, V+A+, A-V+, A+V-; and four nontarget stimulus types: V-, A-, Congruent A-V-, Incongruent A-V-), presentation of these stimulus types were equiprobable, and there was 64 trials with each stimulus type. Therefore, a total of 576 trials were included under each attentional load condition in the experiment. To avoid the fatigue, these trials were divided into 4 main blocks of 144 trials each under each load condition.

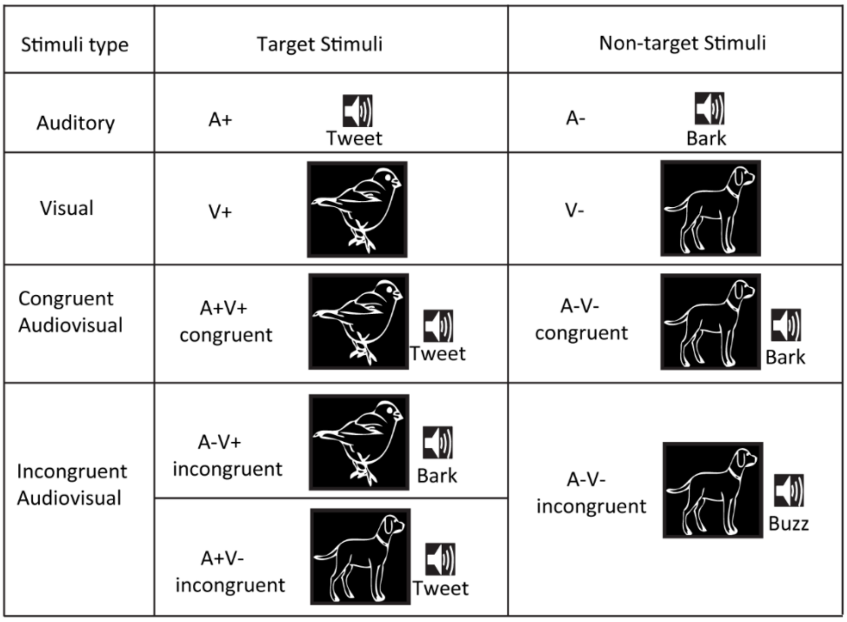

Fig. 1 *The stimulus types used in the animal identification tasks. In this study, five target stimulus types and four non-target stimulus types were derived from the four basic stimulus types. The stimulus under each type is just one example.*

The stimuli in the RSVP task consisted of 23 distractor letters of the alphabet (A, C, D, E, F, J, H, J, K, L, M, N, P, Q, R, S, T, U, V, W, X, Y, Z) and seven digits (2, 3, 4, 5, 6, 7, 9). Some letters (I, B, O) and digits (1, 8, 0) did not appear in the RSVP streams because the visual similarity between the letters and digits could be confusing to the participants. The RSVP streams were presented continuously during the animal identification task (Fig. 2). Each letter/number (subtending 2.0˚ × 2.0˚) in the RSVP stream was presented centrally for 146 ms.

2.2.3 Experimental design and procedure

In the present study, we employed a dual-task design to explore whether semantic congruency modulates the effects of attentional load on AV integration. First, we controlled for semantic congruency in the AV integration task; the semantically congruent/incongruent stimuli comprised animal pictures presented along with either congruent or incongruent auditory stimuli. Second, we adopted the rapid serial visual presentation (RSVP) task used in Gibney et al (2017) as the distractor task to impose different levels of attentional load as follows: no load, low load, and high load [30]. Specifically, the participants simultaneously performed the AV integration task and a distractor task that required them to search a central RSVP stream for either a yellow letter (low load) or a white digit (high load). In addition, under the no-load condition, the participants were instructed to ignore the presented RSVP stream. Additionally, previous dual task studies have utilized similar RSVP streams composed of letters and numbers with a colour change representing a low load target and/or a number representing a high load target [13.14,30].

Our study included three attentional load condition types by adopting an RSVP task, namely, no load, low load, and high load. Under the no-load condition, although the pictures and sounds of animals were presented simultaneously with RSVP streams, participants simply needed to perform the animal identification task (participants had to judge whether the present animal image or sound is a “bird”) and were not instructed to search for the targets in the RSVP streams. In our experiments, each trial began with a 400-ms presentation of the fixation cross to indicate the beginning of a new trial. An animal picture was randomly presented alongside the first through fifth letter of the RSVP stream in each trial. During the experiment, participants were instructed to make a button-press response (the “F” button on the computer keyboard) as soon as possible with their right index finger when a picture or sound target (“bird”) occurred. A blank interface (1000 ms) was presented to ensure sufficient time to respond to the animal identification task (Fig. 2).

The low-load condition consisted of the presentation of an RSVP yellow letter detection task and the animal identification task [13,14,30]. In the animal identification task, the stimuli and procedures were identical to those under the no-load condition (participants were asked to judge whether the image or sound of an animal is a “bird”), while in the RSVP task, the participants were required to detect infrequent yellow letters. Each trial began with a central fixation cross presented for 400 ms, followed by a stream of seven characters (letters or numbers), which were continuously displayed at a rate of 6 Hz. Specifically, these different letters were sequentially presented, being randomly replaced every 146 ms. This random replacement was restricted in such a way that a letter was always replaced with a different letter or digit. The target of the RSVP task was presented with equal probability in the first through seventh positions in the stream. The letters in the stream were chosen randomly prior to each trial, with the sole restriction being that no distractor was repeated within a given stream. Specifically, the RSVP streams in each trial had a 25% probability of containing no numbers or yellow letters, a yellow letter only, a number only, or a yellow letter and a number, thus resulting in a 50% probability of a target being present in each trial for all attentional-load conditions. With respect to the RSVP task, participants were asked to respond at the end of each trial, i.e., after the red fixation point (1000 ms) appeared, subjects were asked to press the “J” button if they observed a target during the RSVP task (Fig. 2).

Under the high-load condition, while the target of the RSVP task was a digit, the other requirements were the same as those under the low-load condition, notably, because the task of searching for digits in a series of letters (high load) requires a higher level of semantic processing and more attentional resources than the task of searching only for a specific colour under the low-load condition [13,14,30]. In this way, by increasing the difficulty of the distractor task, we can control the attentional resource that can be utilized by audio-visual integration processing.

The experiment included 4 blocks of 144 trials each under each load condition, and each block lasted approximately 7 min. Thus, it takes about 28 min for each load condition. Participants were permitted to take breaks between blocks. In addition, each load condition was completed in a separate block, and the order in which participants completed the load condition blocks was randomized and counterbalanced across participants. Before the experiment was officially started, all participants engaged in a practice experiment with 30 trials to ensure that they correctly understood the experimental procedures and responded correctly to the different tasks.

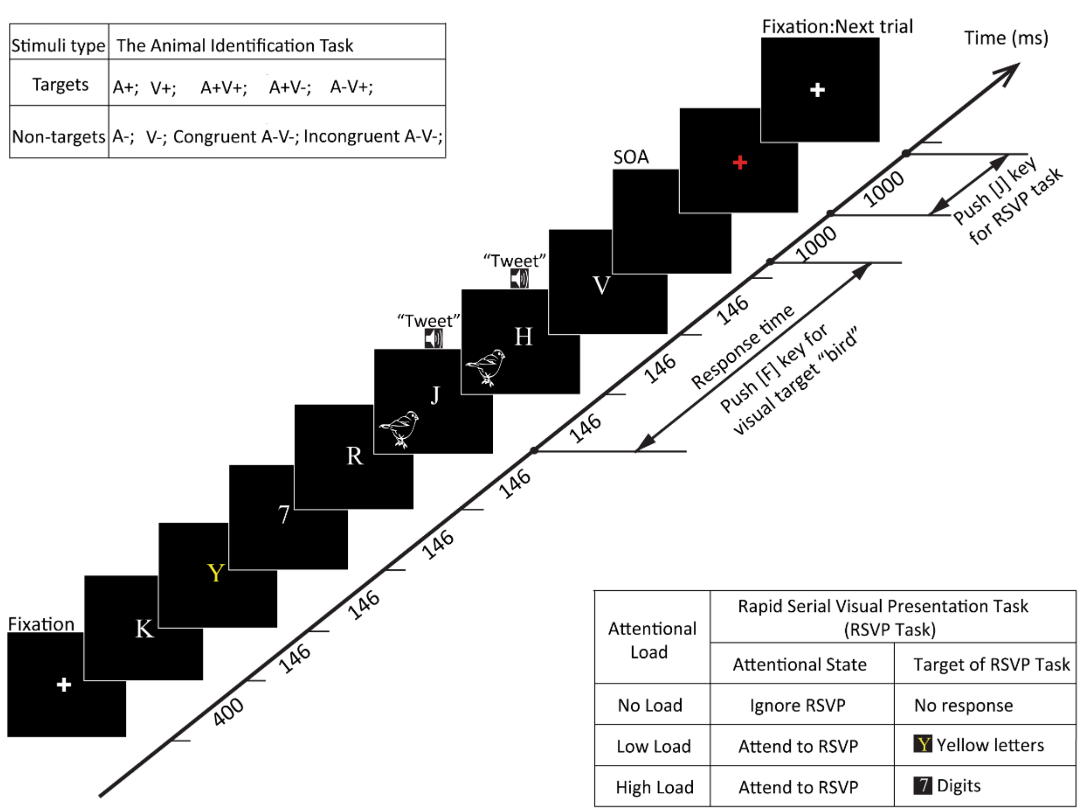

Fig. 2 *A schematic representation trial in which both the animal identification task and the RSVP task were run simultaneously. The participants must judge whether the presented animal image or sound represents a "bird" while ignoring the RSVP streams (no load), reporting yellow letters (low load), or reporting numbers (high load). Each trial began with a central fixation cross (400 ms), followed by a stream of seven characters (letters or numbers), which were sequentially presented with random replacement every 146 ms, while an animal picture or sound (300 ms) was randomly presented alongside the first to fifth letter of the RSVP streams. Participants should respond as soon as possible to the "bird" picture or sound by pressing the "F" key, and they were asked to press the "J" key for the target of the RSVP task when the red fixation point appeared (1000 ms).*

2.2.4 Data analysis

Because Bayesian analysis provides a measure of evidence regarding how much more probable the null hypothesis is compared with the alternative hypothesis [38] and does not depend on the stopping rule [39], for all tests, in addition to p-values, Bayes factors are also reported. A Bayes factor above 3 is indicative of substantial evidence for the alternative hypothesis, whereas a Bayes factor below 1/3 indicates substantial evidence for a null hypothesis; between these values indicates the data are insensitive [39]. Bayes factors were calculated using a half-normal distribution. In addition, in each analysis, the degrees of freedom were corrected using the Greenhouse-Geisser correction when the Mauchly's test indicated that the assumption of sphericity had been violated.

(1) Analysis of the influence of the distractor task

First, to check the RSVP performance to verify that participants accurately performed the distractor task (because they could have simply ignored it and only attended the primary task), the percentage of accuracy under different load conditions were analysed. A Shapiro-Wilk test was conducted to confirm the assumption of a normal distribution in low-load and high-load conditions. If the Shapiro-Wilk test was not significant, the repeated-measures ANOVA for comparisons between different load conditions were conducted. If the Shapiro-Wilk test was significant, we used the one-way nonparametric repeated-measures analyses of variance (ANOVA; the Friedman test) for comparisons. $P<0.05$ was considered to be statistically significant.

Second, we calculated the relative performance under the no load, low load and high-load conditions for all stimuli (A+, V+, A+V+, A+V-, A-V+) to explore whether attentional load significantly disrupted the RTs for the AV integration task.

Additionally, dual-task interference was quantified by calculating a dual task effect (DTE) of each task [26]. To test whether the load manipulation worked, we calculated the DTE of the changes in response time in the multisensory task between the dual task and single task to compare the trial types. For the variables in which higher values indicate worse performance (e.g., response time [RT]), the DTE was calculated as follows:

$$DTE(\%) = \frac{-(\text{dual task RT} - \text{single task RT})}{\text{single task RT}} \times 100\% \quad (1)$$

[26]. Similar measures have been used in other published dual task paradigm studies [30]. Therefore, negative DTE values indicate that attentional load decreased performance (i.e., dual-task cost). We calculated the DTE of the changes in the response time between the no-load and low-load conditions (but not between the no load and high-load conditions) to compare the trial types. We conducted a repeated-measure analysis of variance with DTE as the dependent factor and stimuli modalities (A+, V+, A+V+, A+V-, and A-V+) as the independent factor to explore whether the attentional load has different influences on different stimuli modalities.

(2) Analysis of the audiovisual integration task

Response times (RTs) are defined as the times between the onset of the target presentation and the behavioural response. Incorrect trials and trials with response times shorter than 200 ms or longer than 1200 ms were also excluded from the analysis (3.11%). Median RTs, accuracy, and response distributions for each trial type were calculated for each subject. The median response times of each participant under each condition were used in the response time analysis as response time distributions are generally skewed and the median is less affected by the presence of outliers. Median RTs were calculated for attentional-load conditions, i.e., no load, low load, and high load, and were separated by modality, i.e., V+, A+, A+V+, A+V-, and A-V+. The main effects and interactions of load condition and modality type were analysed using repeated-measures ANOVA with 5 stimuli modalities (V+, A+, A+V+, A+V-, A-V+) * 3 attentional loads (no load, low load, high load). Percentage of accuracy was analyzed by nonparametric repeated-measures analyses of variance (ANOVA;

the Friedman test). $P<0.05$ was considered to be statistically significant.

(3) Calculation of cumulative distribution functions

Race model of semantically congruent AV stimuli To test whether participants integrated the semantically congruent AV stimuli under each load [40-41], we used the individual cumulative distribution functions (CDFs) of each target modality in each load condition to calculate the race model using the following formulas:

$P(RT_{\text{Race model}} < t) = P(RT_A < t) + P(RT_V < t)$.

This Inequality does not require the channel processing times to be stochastically independent, and this prediction allows one to rule out all separate-activation models [41-42]. Thus, it is suitable for calculating the race model inequality. In this formula, the race model provides the probability (P) of a RT that is less than a given time in milliseconds, where time ranges from 200-1200 ms after stimulus onset. Additionally, race model inequality violation is based on the combination of the unimodal auditory and unimodal visual CDFs [41]. The percentiles of the semantically congruent audiovisual CDF of each participant in each load condition were compared to the corresponding race model CDF (e.g., no load AV CDF vs. no load race model CDF) at each time bin to test for race model inequality violations [42]. Two-tailed paired *t*-tests were used to analyse race model inequality violations [43] (the resulting p-values were Bonferroni corrected; $p < 0.05$). Significant violations of the race model (i.e., $RT_{AV} < RT_{\text{Race model}}$) indicate audiovisual interactions that exceed statistical facilitation.

Because each subject has a different time course for his or her responses, averaging difference curves across individuals may not provide a complete indication of group differences [44]. Moreover, in previous studies, the positive area under the difference curve was used as a measure of audiovisual integration, and it was not affected by timing differences across individuals [45]. Thus, we specifically calculated the positive area under the difference curve (i.e., the difference in probability of the congruent AV CDF and the race model CDF for the RT range from 200 to 1200 ms) to test for differences in race model inequality violation between different attentional loads. We also followed the approach described with RSE-box to analyse the positive area under the difference curve [46]. To extract the positive area under the difference curve, all negative probabilities (no race model violation) were set to a value of zero, and only the positive area under the curve was calculated for all participants [43-44]. We then compared the positive area under the difference curve between attentional-load conditions using a repeated-measure ANOVA with the factor attentional load (no load, low load, and high load) to explore how attentional loads influence the integration of semantically congruent AV stimuli.

(4) Distractor effect of semantically incongruent AV stimuli

To assess the distractor effect of semantically incongruent AV stimuli, the CDFs for responses to unisensory targets were subtracted from the CDFs for responses to incongruent AV targets, yielding a relative distractor effect [5]. Specifically, the CDFs for responses to unisensory auditory targets (A+) were subtracted from the CDFs for responses to incongruent AV targets (A+V-; auditory targets with visual distractors) to obtain a measure of the visual distractor effects for incongruent AV targets; the comparison between unisensory visual targets (V+) and incongruent AV targets (A-V+; visual targets with auditory distractors) produced the auditory distractor effect for incongruent AV targets. At each time bin, we performed two-tailed paired *t*-tests to evaluate the difference in probability between unisensory CDFs and incongruent AV CDFs from 200 to 1200 ms in each load condition to assess significant differences in the visual/auditory distractor effect in different load conditions ($p < 0.05$; the resulting p-values were Bonferroni corrected).

Furthermore, we specifically calculated the negative area under the difference curve (i.e., the difference in probability of the unimodal CDF and the incongruent AV CDF for the RT range from 200 to 1200 ms) to examine differences in the distractor effect for incongruent AV targets in different load conditions. To extract the negative area under the difference curve, all positive probabilities were set to a value of zero, and only the negative area under the curve was calculated for all participants. We determined the negative area under the difference curve by calculating the trapezoidal area between each time bin that produced a negative distractor effect. Each trapezoidal negative area between each time bin was summed to provide a total negative area for each load condition. We then compared the negative area under the difference curve between different attentional-load conditions using a repeated-measure ANOVA with the factors of attentional load (no load, low load, high load) to determine how the visual/ auditory distractor effect for AV incongruent targets was influenced by attentional-load conditions.

2.3 Results

2.3.1 Response time and accuracy

We performed two planned comparisons to study (1) RT bimodal facilitation (A+V+ compared to A+ and V+ together) under all load conditions and (2) the distractor effect (comparison of V+ with A-V+ and A+ with A+V-) under different attentional loads. To determine how attentional load interacts with semantic congruency to influence audiovisual integration, we conducted repeated-measures ANOVA on median response time using stimulus modality (V+, A+, A+V+, A-V+, A+V-) and attentional load (no load, low load, high load) as factors. Significant main effects of stimulus modality [$F(1.382, 23.498) = 44.798$,

$p < 0.001$, $\eta^2 = 0.725$, $BF_{(10)} = 6.876 \times 10^{28}$] and load [$F(2,34) = 37.744$, $p < 0.001$, $\eta^2 = 0.689$, $BF_{(10)} = 1.33 \times 10^{33}$] were observed. However, we did not find a significant interaction between stimulus modality and load [$F(3.02,51.32) = 1.789$, $p = 0.084$, $\eta^2 = 0.095$, $BF_{(10)} = 0.034$]. To test our main hypotheses in detail, we then analysed this result separately under different load conditions by conducting Plan-tests. Post hoc subsidiary analyses with Bonferroni adjustment for multiple comparisons (plan-tests) demonstrated the following.

1) The median RTs for the A+V+ trials were significantly faster than those for the V+ trials [no load: $t(17) = 14.6$, $p < 0.001$, $BF_{(10)} = 5.25 \times 10^{8}$; low load: $t(17) = 10.25$, $p < 0.001$, $BF_{(10)} = 8.71 \times 10^{5}$; high load: $t(17) = 9.67$, $p < 0.001$, $BF_{(10)} = 3.08 \times 10^{5}$] or the A+ trials under each load condition [no load: $t(17) = 4.5$, $p = 0.004$, $BF_{(10)} = 85.28$; low load: $t(17) = 6.78$, $p < 0.001$, $BF_{(10)} = 6.972 \times 10^{3}$; high load: $t(17) = 8.5$, $p = 0.006$, $BF_{(10)} = 1.79 \times 10^{5}$] (Fig. 3). This finding suggests that the identified speed advantage for the semantically congruent AV target over both types of unisensory targets was observed under all load conditions.

2) The median RTs for the A+V- trials were not significantly slower than those for the A+ trials under all load conditions [no load: $t(17) = 2.25$, $p = 0.413$, $BF_{(10)} = 1.676$; low load: $t(17) = 0.67$, $p = 0.529$, $BF_{(10)} = 0.292$; high load: $t(17) = 1.29$, $p = 0.209$, $BF_{(10)} = 0.506$] (Fig. 4a). In addition, the median RTs for the A-V+ trials were significantly slower than those for the V+ trials under the no-load condition [$t(17) = 3.67$, $p = 0.008$, $BF_{(10)} = 43.5$], but there was no significant difference under the low-load and high-load conditions [low load: $t(17) = 2.67$, $p = 0.291$, $BF_{(10)} = 2.213$; high load: $t(17) = 1.0$, $p = 0.313$, $BF_{(10)} = 0.389$] (Fig. 4b). This observation revealed an auditory interference effect only under the no-load condition, and attentional load hindered this distractor effect.

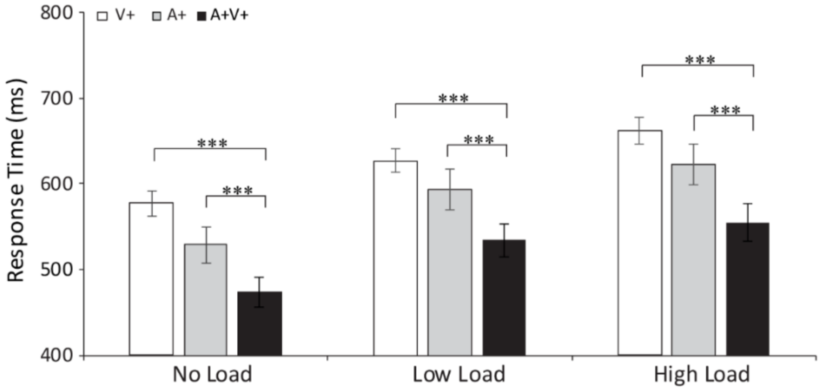

Fig. 3 *The median response times in the animal identification task. Comparison of the magnitudes of the mean response times in the unisensory visual (V+), auditory (A+), and bimodal congruent (A+V+) trials under the no-load, low-load, and high-load conditions. Error bars represent the standard errors of the means. ***p < 0.001, **p < 0.01.*

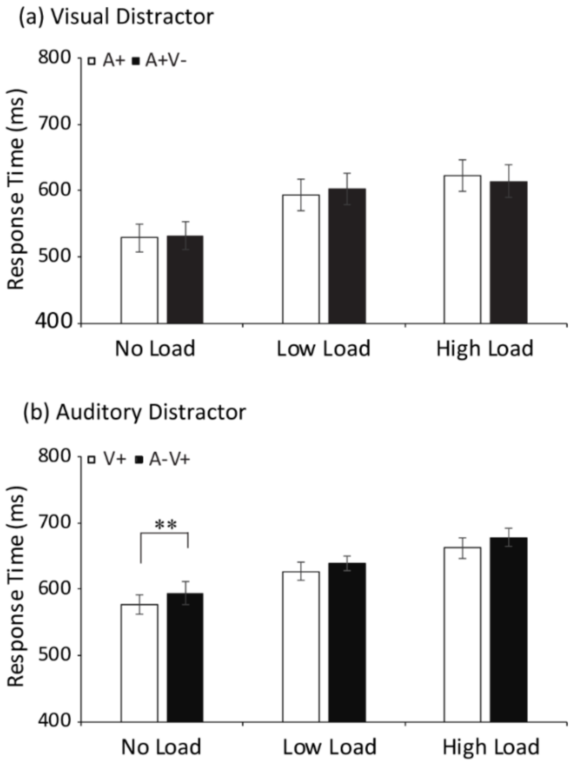

Fig. 4 *The median response times in the animal identification task are presented. (a) Comparison of the magnitudes of median response times for unimodal auditory trials (A+) and bimodal incongruent A+V- trials under no-load, low-load, and high-load conditions. (b) Comparison of the magnitude of median response times for unisensory visual trials (V+) and bimodal incongruent A-V+ trials under no-load, low-load, and high-load conditions. **p < 0.01.*

The accuracy in the AV integration task in all load conditions violated the Shapiro-Wilk tests (all $W < 1$, all $p < 0.01$), the non-parametric Friedman tests on the accuracy of AV integration task showed significant differences under different load conditions ($\chi^2(14) = 77.22$, p < 0.001). The Wilcoxon signed-rank tests on the coefficient of variance showed significant influences for some stimulus types in the no-load condition (V+ - V+A+, W(18)= -2.414, p= 0.016A+ - V+A+, W(18)= -2.512, p= 0.012), low-load condition (V+ - V+A+, W(18)= -2.99, p= 0.003 ; A+ - V+A+, W(18)= -2.61, p= 0.009) and high-load condition (V+ - V+A+, W(18)= -2.95, p= 0.003 ; A+ - V+A+, W(18)= -3.308, p= 0.001). However, there was no significant difference for other stimulus types in the no-load condition (V+ - V+A-, W(18)= -0.061, p= 0.952; A+ - V-A+, W(18)= -0.71, p= 0.944), the low-load condition (V+ - V+A-, W(18)= -1.85, p= 0.065; A+ - V-A+, W(18)= -0.284, p= 0.776) and the high-load condition (V+ - V+A-, W(18)= -1.51, p= 0.131; A+ - V-A+, W(18)= -1.62, p= 0.106). These results showed that although advantageous nature of A+V+ stimuli over V+ and A+ were observed under different load conditions, the distracting nature of A-V+ and A+V- stimuli was not found under all load conditions.

Table 1 *Median accuracy (%) and response times (RTs, ms) with standard deviations (SDs) for each trial type under no load, low load, and high-load conditions.*

	No Load		**Low Load**		**High Load**	
	RTs (SD)	**Accuracy (SD)**	**RTs (SD)**	**Accuracy (SD)**	**RTs (SD)**	**Accuracy (SD)**
V+	577.2 (61.2)	99.0 (1.4)	626.9 (56.6)	96.2 (4.1)	662.2 (63.6)	96.1 (5.9)
A+	528.3 (91.5)	97.3 (4.6)	593.4 (102.5)	95.1 (7.0)	622.2 (101.8)	92.5 (9.2)
A+V+	473.2 (76.3)	99.9 (0.4)	534.5 (81.7)	93.3 (1.4)	554.7 (93.7)	98.9 (3.0)
A-V+	593.9 (72.2)	98.9 (2.1)	639.4 (48.6)	98.0 (3.2)	678.4 (59.3)	98.4 (3.3)
A+V-	531.2 (88.7)	97.5 (4.0)	602.1 (102.0)	95.1 (5.4)	613.9 (103.6)	95.1 (6.9)

2.3.2 Race model violation of semantically congruent AV stimuli

Consistent with the median RT comparisons that showed similar significant multisensory gains under all attentional-load conditions, the comparisons between the semantically congruent AV CDF and the race model CDF under each load condition for each time bin revealed significant race model inequality violations for all load conditions ($p < 0.05$, paired t-test, 2-tailed, Bonferroni-corrected). A significant race model inequality violation was observed from 430 ms to 500 ms in the no-load condition ($p < 0.05$), from 410 ms to 520 ms in the low-load condition ($p < 0.05$), and from 480 ms to 540 ms in the high-load condition ($p < 0.05$). The range of RTs in which the significant race model inequality violation was observed under the no-load condition was not greater than that observed for the low-load and high-load conditions.

In addition, the positive area under the curve was compared between different load conditions (Fig. 5). The repeated-measures ANOVA revealed that attentional load did not significantly modulate the positive area under the curve [$F(1.747, 29.69) = 0.635$, $p = 0.517$, $\eta^2 = 0.036$, $BF_{(10)} = 0.231$]. Notably, a Bayes factor below 1/3 indicates substantial evidence for a null hypothesis (Dienes, 2014), and hence, the Bayesian analyses of the positive area under the curve between different load conditions clearly showed evidence for no effect of attentional load on the positive area under the curve. The post hoc paired t-tests (Bonferroni-corrected) revealed that the positive area under the curve in the no-load condition ($M = 17.17$ ms, $SE = 2.57$) was also not significantly larger than that in the low-load condition [$M = 14.94$ ms, $SE= 2.28$, $t(17) = 0.907$, $p= 0.377$, $BF_{(10)}= 0.349$] and high-load condition [$M = 18.28$ ms, $SE = 3.08$, $t(17) = -0.321$, $p= 0.752$, $BF_{(10)}= 0.255$]; there was also no significant difference between the low-load and high-load conditions [low load/high load: $t(17) = -1.096$, $p = 0.288$, $BF_{(10)}= 0.410$] (Fig. 5d). These results indicated that attentional load did not affect the overall strength of semantically congruent AV integration.

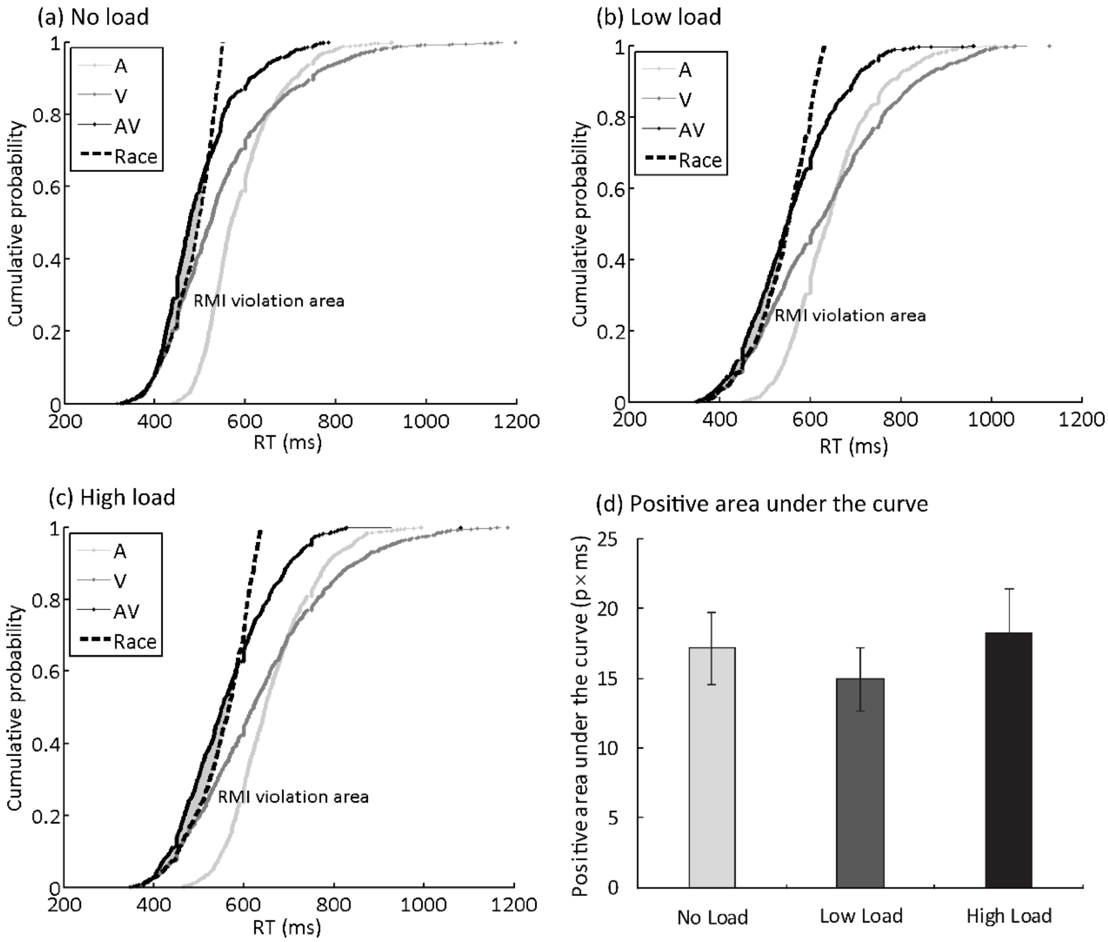

Fig. 5 *Distributions of the response times under different load conditions. (a) Cumulative distribution functions (CDFs) for the discrimination response times to auditory, visual, semantically congruent audiovisual stimuli, and race model under no-load condition. (b) CDFs under the low-load condition. (c) CDFs under the high-load condition. (d) No significant difference was observed across different load conditions for the positive area.*

Table 2 *Peak Benefit (%), Peak Latency (ms), and Time Window (ms) of Semantically Congruent AV Integration in each Load Condition.*

	Semantically Congruent AV Integration		
	Peak benefit (%)	Peak latency (ms)	Time window (ms)
No Load	13.74	480	430 - 500
Low Load	7.75	490	410 - 520
High Load	7.65	510	480 - 550

2.3.3 The interference effect produced by semantically incongruent AV stimuli

To assess the effects of non-matching cross-modal distractors, we compared the response distributions for different unisensory trials with the response distributions for non-matching multisensory trials under different attentional conditions (Fig. 6).

Visual distractor effect A comparison between the auditory (A+) CDF and semantically incongruent A+V- CDF in each time bin showed a visual distractor effect; however, this visual distractor effect was only observed under the no-load condition ($p < 0.05$, paired t-tests, 2-tailed, Bonferroni-corrected). Specifically, a visual distractor effect was observed at 770-930 ms in the no-load condition ($p < 0.05$), but no visual distractor effect was found under the low-load or high-load condition. The negative area under the curve was compared between the different load conditions (Fig. 8b). The repeated-measures ANOVA revealed a main effect of load [$F(1.285, 21.84) = 9.235$, $p = 0.004$, $\eta^2 = 0.352$, $BF_{(10)} = 195.3$]. The post-hoc test revealed that the negative area under the curve of the visual distractor effect in the no-load condition (M = -10.95 ms, SE =2.8) was significantly larger compared with the low-load condition [M = -1.18 ms, SE =0.82; $t(17) = 3.72$, $p = 0.005$, $BF_{(10)} = 23.7$] and high-load condition [M = -2.76 ms, SE =0.88; $t(17) = 2.69$, $p = 0.046$, $BF_{(10)} = 3.7$], but there was no difference in the negative area between the low-load and high-load condition [$t(17) = 1.2$, $p = 0.74$, $BF_{(10)} = 0.45$]. This result suggested that attentional load reduced the visual distractor effect.

Auditory distractor effect In addition, the comparison between visual (V+) CDF and semantically incongruent A-V+ CDF in each time bin revealed an auditory distractor effect, but this auditory distractor effect was only observed under the no load and low-load conditions ($p < 0.05$, paired t-tests, 2-tailed, Bonferroni-corrected, Fig. 6c). Specifically, an auditory distractor effect occurred at 520-660 ms under the no-load condition and at 610-760 ms under the low-load condition; the auditory distractor effect was not found under the high-load condition. The negative area under the curve was compared between the different load conditions (Fig. 6d). The repeated-measures ANOVA revealed a main effect of load [$F(1.078,18.323) = 12.168$, $p = 0.002$, $\eta^2 = 0.417$, $BF_{(10)} = 1194.9$]. The post-hoc test showed that the negative area of the auditory distractor effect was significantly larger in the no-load condition (M = -15.95 ms, SE =3) as compared with the high-load condition [M = -2.96 ms, SE =0.72; $t(17) = 4.142$, $p = 0.002$, $BF_{(10)} = 52.05$], but not compared with the low-load condition [M = -8.4 ms, SE =0.84; $t(17) = 2.32$, $p = 0.099$, $BF_{(10)} = 1.992$]. The positive area under the curve in the low-load condition was significantly larger than the high-load condition [$t(17) = 7.39$, $p < 0.001$, $BF_{(10)} = 1.72\times10^4$]. This result suggested that attentional load reduced the auditory distractor effect.

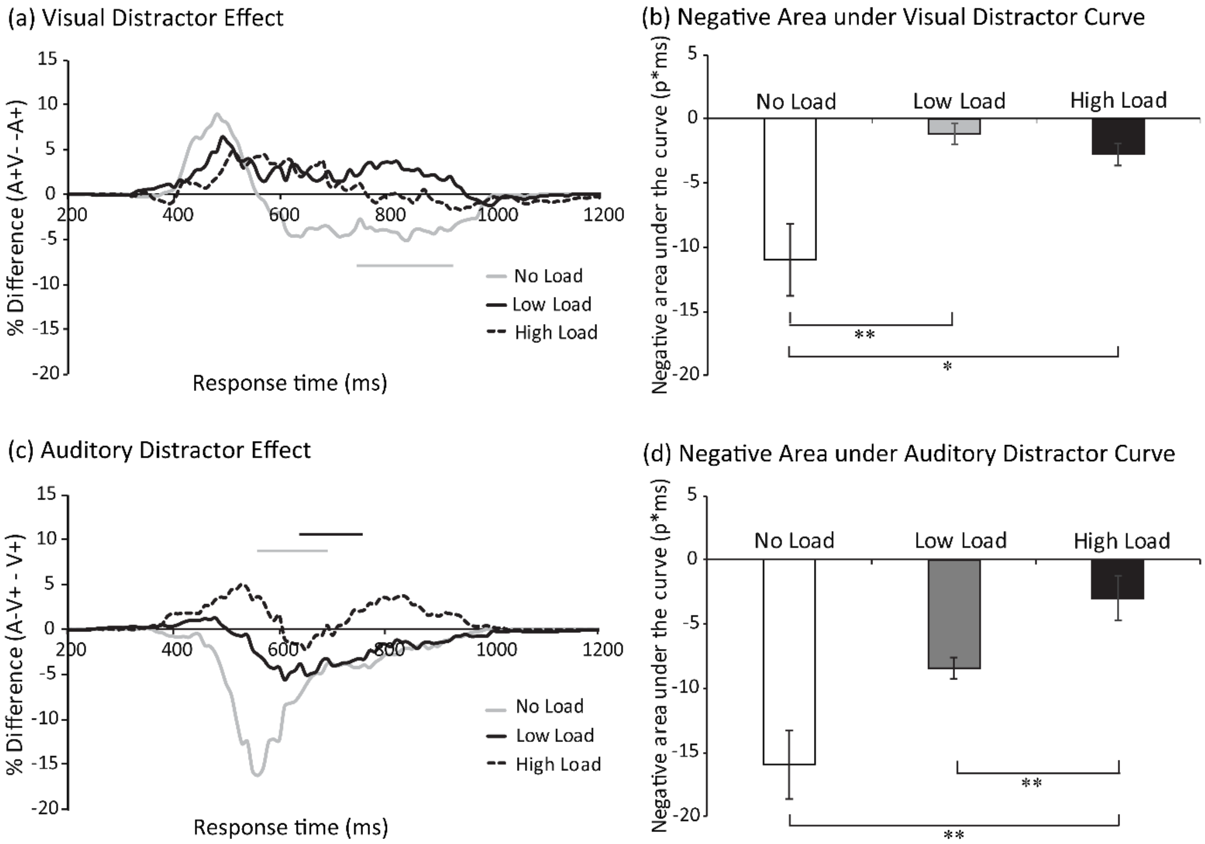

Fig. 6 *(a) Visual and auditory distractor effects under no-load, low-load, and high-load conditions. The subtraction of A+ CDF from A+V- CDF yields the visual distractor effect, but a significant visual distractor effect is only present under the no-load condition. (c) The subtraction of V+ CDF from A-V+ CDF yields the auditory distractor effect, but a significant auditory distractor effect is only present under the no-load and low-load conditions. The average negative area under the curve in each load condition was plotted separately for the visual distractor (b) and auditory distractor (d) effects, demonstrating that both were reduced by attentional load. **p < 0.01, *p < 0.05.*

Table 3 *Peak Benefit (%), Peak Latency (ms), and Time Window (ms) of different Distractor Effects in each Load Condition.*

	Visual Distractor			Auditory Distractor		
	Peak benefit	Peak latency	Time window	Peak benefit	Peak latency	Time window
No Load	-5.12	840	770-930	-16.3	560	520-660
Low Load	-1.22	990	--	-5.53	610	610 - 760
High Load	-1.70	940	--	-2.26	650	--

2.3.4 The influence of the distractor task

First, because the Shapiro-Wilk test for the accuracy of the RSVP task under each load condition was not significant (low load: $W=0.948$,$p=0.388$; high load: $W=0.931$, $p=0.202$), we conducted the repeated-measures ANOVA to determine whether accuracy in the RSVP task was reduced by attentional load. The results indicated that the accuracy of the RSVP task was significantly higher under the low-load condition ($M = 90.4\%$, $SE = 0.79$) than that under the high-load condition ($M = 84.8\%$, $SE = 1.37$) [$F(1,17) = 23.84$, $p < 0.001$, $\eta^2 = 0.584$, $BF_{(10)}= 340.0$]. Moreover, the accuracy of the RSVP performance was above 80%, indicating that the participants accurately performed the distractor task; the participants did not only perform the AV integration task under the LL and HL conditions.

Second, the repeated-measures ANOVA using stimulus modality (V+, A+, A+V+, A-V+, A+V-) and attentional load (no load, low load, high load) as factors in the AV integration task revealed a main effect of load [$F(2,34) = 37.744$, $p < 0.001$, $\eta^2 = 0.689$, $BF_{(10)} = 1.33\times10^{33}$], the post-hoc test showed that inter-participant median RTs for the AV integration task were significantly slower under the low load ($M = 617$, $SE = 18$) compared with the no load ($M = 557$, $SE = 17$, $t(17) = 2.78$, $p = 0.034$, $BF_{(10)} = 1.29\times10^{4}$) condition, and the median RTs under the high load ($M = 642$, $SE = 19$) were slower than those under the low load ($t(17) = 6.67$, $p < 0.001$, $BF_{(10)} = 1.45\times10^{16}$) condition. These results indicated that the high load task was more demanding. Furthermore, the attentional load significantly disrupted the RTs, regardless of the sensory modality (Fig. 7), specifically for V+ stimuli [no load/low load: $t(17) = -5.6$, $p < 0.001$, $BF_{(10)} = 854.35$; no load/high load: $t(17) = -8.09$, $p < 0.001$, $BF_{(10)} = 5.2\times10^{4}$; low load/high load: $t(17) = -4.71$, $p = 0.001$, $BF_{(10)} = 1.27\times10^{2}$], A+ stimuli [no load/low load: $t(17) = -6.0$, $p < 0.001$, $BF_{(10)} = 1.7\times10^{3}$; no load/high load: $t(17) = -6.0$, $p < 0.001$, $BF_{(10)} = 1.48\times10^{3}$; low load/high load: $t(17) = -1.46$, $p = 0.18$, $BF_{(10)} = 0.558$], A+V+ stimuli [no load/low load: $t(17) = -3.5$, $p < 0.001$, $BF_{(10)} = 1.3\times10^{3}$; no load/high load: $t(17) = -6.42$, $p < 0.001$, $BF_{(10)} = 3.25\times10^{3}$; low load/high load: $t(17) = -2.9$, $p = 0.036$, $BF_{(10)} = 4.57$], A-V+ stimuli [no load/low load: $t(17) = -3.5$, $p = 0.009$, $BF_{(10)} = 15.05$; no load/high load: $t(17) = -7.91$, $p < 0.001$, $BF_{(10)} = 455.1$; low load/high load: $t(17) = -4.14$, $p = 0.003$, $BF_{(10)} = 38.55$] and A+V- stimuli [no load/low load: $t(17) = -5.0$, $p < 0.001$, $BF_{(10)} = 243.3$; no load/high load: $t(17) = -4.94$, $p < 0.001$, $BF_{(10)} = 185.2$; low load/high load: $t(17) = -1.08$, $p = 0.31$, $BF_{(10)} = 0.389$]. In summary, the response times to all stimulus modalities (A+, V+, A+V+, A+V-, A-V+) were significantly slower under high-load than under no-load (all $F > 1$, all $p < 0.01$). Hence, the identification of targets in the AV integration task was slower under low-load and high-load conditions versus no-load conditions regardless of the sensory modality.

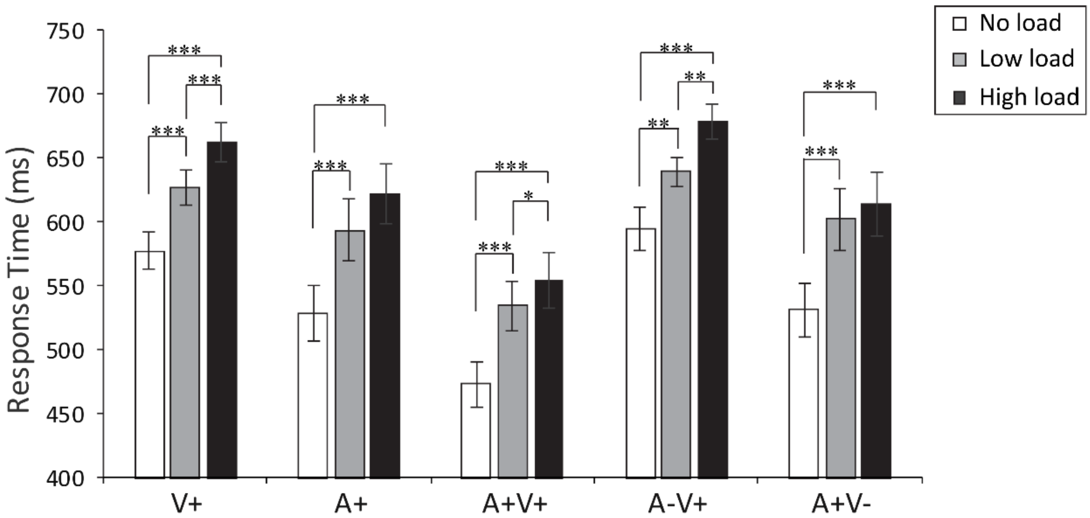

***Fig.** 7 The median RTs under the unimodal (A+ and V+), bimodal congruent (A+V+) and bimodal incongruent (A-V+ and A+V-) conditions are presented under different load conditions. The response times to all stimuli generally increased as the load increased. The error bars represent the standard error of the mean. ***p < 0.001, **p < 0.01, *p < 0.05.*

Additionally, the DTE values of the response time in all trial types in the AV integration task were negative, indicating that attentional load decreased performance (Fig. 8). A repeated-measures ANOVA of the 5 stimulus modalities (V+, A+, A+V+, A-V+, and A+V-) did not show a significant main effect of the stimulus modalities [$F(1.96, 33.32) = 2.98$, $p = 0.065$, $\eta^2 = 0.149$, $BF_{(10)} = 2.3$], suggesting that the DTEs of the changes in response time in the AV integration task did not significantly differ across the trial types.

Overall, these results demonstrated two key findings. First, we checked the RSVP performance and verified that participants accurately performed the task, and the load manipulation was indeed functional (the RSVP performance was lower under the high-load condition than the low-load condition). Second, the load manipulation indeed interfered with the target processing in the AV integration task because the response times to all target stimuli was significantly decreased by attentional loads.

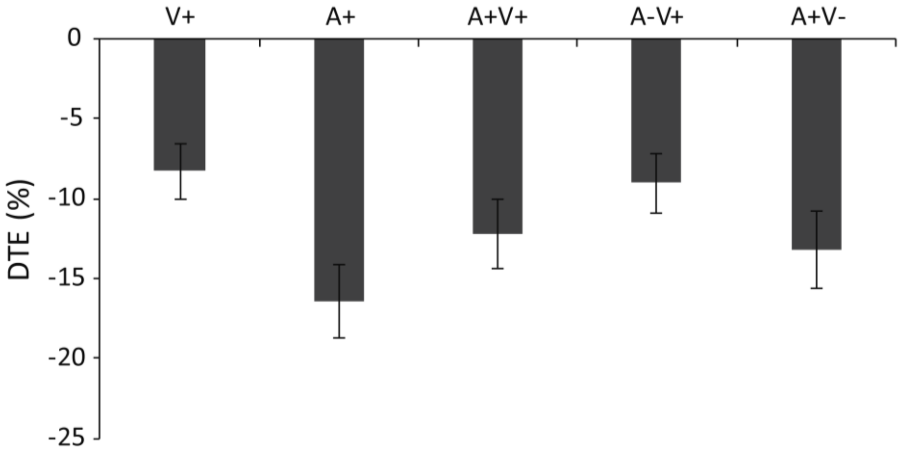

Fig. 8 *The response time DTE of different stimulus types in the identification task. The sign of the DTE for response time was reversed so that the increased response time is represented as a negative DTE.*

2.4 Discussion

The present study sought to determine how attentional load interacts with semantic congruency to influence the AV integration of common objects. We used an RSVP task to manipulate the amount of attentional resources that were available for the integration processing of semantically congruent and incongruent animal sounds and images. Our results revealed that attentional load did not eliminate the AV integration of semantically congruent animal sounds and images (Figures 3 and 5). However, semantically incongruent AV stimuli were not integrated (as there was no multisensory facilitation) under all load conditions, and attentional load attenuated the multisensory interference effect produced by semantically incongruent animal sounds and images (Figures 4 and 6). The integration of semantically congruent AV object features appeared to be more robust to attentional load manipulation than the multisensory interference effect of semantically incongruent AV object features. Thus, our finding provides evidence that semantic congruency modulates the effect of attentional load on the AV integration of common objects.

2.4.1 The effect of attentional load on the facilitation effect of semantically incongruent AV object features

To the best of our knowledge, the present study is the first to demonstrate that attentional load does not eliminate the AV integration of semantically congruent animal sounds and images (Figures 3 and 5). Regarding the facilitation effect produced by semantically congruent bimodal stimuli, it has been proposed that relevant semantic unimodal information could be rapidly integrated into a coherent multisensory representation (i.e., within 100 ms in certain cases)[1,16], and that the effective mental representation formed by semantically congruent AV stimulus can be well-matched with the inherent characteristics already present in memory systems [33-35]; thus, the consolidation and integration processing of semantically congruent AV information is enhanced. One possibility that could explain why the integration of semantically congruent AV object features can resist external interference is related to the "attentional load theory", which postulates that tasks involving a high perceptual load that requires full capacity leave little capacity for the processing of irrelevant distractor information [22,23]. However, as the goal of the present study involves identifying a visually presented animal image (or identify an animal sound), the presentation of a semantically congruent animal sound (or a congruent animal image) could provide coherent and useful information for identification of the target; furthermore, most task-relevant inputs can be prioritized given that they are highly relevant to the current task. Therefore, it is difficult for attentional load to hinder the integration of semantically congruent AV object features.

It is noteworthy that one neuroimaging study has explained why multisensory cues retain their ability to capture a participant's attention, even under conditions of attentional or memory load [47-48]. Specifically, the multisensory control may still mediate modulatory effects from higher-order fronto-parietal regions even when there is a uncoupling between cross-modal effects in the visual cortex and working memory/sustained visuospatial attention, such that multisensory interactions between visual-tactile stimuli seem to be relatively unaffected by manipulations of visual load [47]. Similarly, it has been proposed that a multisensory or supramodal cortical region higher in the information processing hierarchy (e.g., polysensory superior temporal sulcus) might send signals to the unisensory cortices to modulate the processing of the features of common objects, even when some features of a particular object are not explicitly attended [33-34], because the neural representations of features of common objects are likely to be strongly and tightly bound together [2]. This phenomenon may indicate that higher-order multisensory cortical regions can still play an important mediating role in the multisensory interaction between semantically congruent audio-visual features of common objects, even without much attentional resources. Moreover, when the time period between prime-target pairs that share a semantic relationship is shorter than 200 ms, the semantic priming processing for prime-target pairs of the same object is relatively automatic [49]. Therefore, the integration of semantically congruent AV object features can also occur even when attentional resources is exhausted.

2.4.2 The effect of attentional load on the multisensory interference effect of semantically incongruent AV object features

Nevertheless, attentional load has a different effect on the multisensory interference effect produced by semantically incongruent AV object features. Consistent with previous findings [35,50], we observed an auditory distractor effect (incongruent A-V+ compared to unimodal V+) and a visual distractor effect (incongruent A+V- compared to unimodal A+) under the no-load condition (see Figures 6a and 6c), but these interference effects of the semantically incongruent animal sounds and images were attenuated by attentional load (Figures 6b and 6d). It is possible that if the presented AV stimuli are semantically incongruent, the mismatch between the actual sensory input and prediction in the memory system could lead to a major update of the internal model of the mental representation [20]; in such a case, the presence of semantically incongruent AV objects could cause a certain degree of an interference effect and impair behavioural performance. Furthermore, the brain does not absorb the mismatched auditory information into the memory system (i.e., it should be rapidly forgotten) if the presented sound is not semantically consistent with the representation of the target images because this incongruent information is useless in the relevant task [51]. Thus, under the conditions of limited and absent attentional resources, the top-down modulatory mechanism underlying selective attention processes may automatically filter task-irrelevant mismatched information, further preventing irrelevant stimuli from entering the memory system, increasing the speed of the forgetting process and resulting in reduced interference effects.

We further observed an asymmetric cross-modal interference effect supporting the visual dominance hypothesis; specifically, the auditory distractor effect (unimodal V+ compared to incongruent A-V+) was stronger than the visual distractor effect (unimodal A+ compared to incongruent A+V-) under all attentional-load conditions (see Figures 4 and 6). When no attentional load is added, it has been proposed that the different interference effects produced by semantically incongruent AV (A+V-, A-V+) stimuli may occur because the attention system itself is not completely supramodal [52]; in other words, attentional modulation of sensory neural processing in the visual cortex can occur at least partially independently from similar attentional modulations to auditory processing [53], and a possible asymmetry may exist in the attentional filtering of irrelevant auditory and visual information [35]. Therefore, during the processing of semantically incongruent AV stimuli, the ability to filter irrelevant visual distractors is stronger compared with irrelevant auditory distractors, resulting in the auditory distractor effect (unimodal A+ compared with incongruent A+V-) is stronger than the visual distractor effect (unimodal A+ compared with incongruent A+V-). Notably, one possibility to consider regarding the asymmetric cross-modal interference effect under increased load conditions is, because studies investigating object-based attention tasks across sensory modalities suggest that attentional resources are at least partially distinct for the visual and auditory sensory modalities [52,53], and the presence of a visual RSVP stream might make participants to focus strongly on the visual modality and occupy a large amount of visual attentional resources, more attentional resources can remain to process task-irrelevant auditory distractors than irrelevant visual distractors. Thus, the auditory distractor effect (A-V+ compared to unimodal V+) will be stronger than the visual distractor effect (A+V- compared to unimodal A+) even under low and high-load conditions.

2.4.3 Research limitations and prospects

One could argue that the RSVP task applied herein (low vs. high load) is not a load manipulation but rather a task switch (colour vs. digit detection task) that interferes with AV integration. We think this possibility exists, as this switch between two tasks in response mappings does cause some interference. Notably, the colour or digit detection task inevitably consumes certain attentional resources and competes for the cognitive resources of AV integration task given that the accuracy of the RSVP task was above 90%, and the performance on the RSVP task decreased as the load increased. Furthermore, the levels of the load manipulation (colour vs. digit detection task) tap into the same type of processing resources since the detection of colours and digits belongs to object recognition (the so-called "what") [54-55], and notably, the task of searching for digits in a series of letters (high load) requires a higher level of semantic processing and more attentional resources than the task of searching only for a specific colour under the low-load condition.

Furthermore, one could also argue that attentional load manipulation (by adopting RSVP tasks) may only interfere with processing in the visual sensory modality (in the AV integration task) but has no effect on processing in the auditory sensory modality. Indeed, when applying the dual-task methodology, a general concern is whether the two tasks compete for the same pool of attentional resources or whether multiple resource pools are used to separately address the various cognitive and perceptual aspects of the two tasks [54]. In fact, some researchers have proposed that the recruitment of shared or distinct attentional resources across sensory modalities is partially task-dependent [55-56] and depends on whether the tasks involve object-based attention (e.g., colour or shape), spatial attention (e.g., localization of stimuli), or both [54]. In addition, it has been proposed that in the visual and auditory sensory modalities, if object-based attention tasks are time-critical, shared resources are recruited across the sensory modalities [54]. Because the main task we adopted is an object recognition task and the distractor task (RSVP task) involving searching for either a yellow letter or a white digit is also an object attention task, we considered the RSVP tasks to interfere with target processing in both sensory modalities in a previous study. Moreover, our results showed that the RSVP task not only interfered with target processing in the visual sensory modality but also significantly interfered with target processing in the auditory sensory modality (Figures 7 and 8), further confirming that the RSVP tasks we adopted interfered with target processing in both sensory modalities.

2.5 Conclusions

The experiments described herein indicate that semantic congruency modulates the effect of attentional load on the audiovisual integration of common objects. Specifically, the performance enhancements associated with semantically congruent AV object features are present even when attentional resources are limited; however, semantically incongruent animal sounds and images were not integrated, and attentional loads influenced the multisensory interference effect produced by incongruent AV object features. Furthermore, based on the fact that the attentional statues of some special subjects will also alter greatly due to the increase of age and the development of some diseases, it is interesting to explore whether the decline of attention is an important factor that influence the automatic integration of semantically congruent AV stimuli.

Chapter 3 Whether Attentional Load Influences the Audiovisual Integration of Selective Attention Depends on Semantic Associations

Summary

Although neuronal studies have shown that attention can automatically spread from an attended visual stimulus of common objects to a task-irrelevant, simultaneously presented auditory stimulus, even when the auditory stimuli are instructed to ignored and semantically conflict with the visual targets. However, there is still little knowledge about how limited attentional resource will influences this audio-visual (AV) integration of common objects when attention is selectively focused on visual modality, and whether semantic associations among the multisensory features of common objects will modulate the influence of limited attentional resource on this cross-modal integration. We manipulated the amount of attentional resources by applying a dual-task paradigm, and constructed three attentional load levels (no-load, low-load, high-load) by using an RSVP task. Semantic associations among AV stimuli were composed of animal pictures presented together with either semantically congruent, incongruent or unrelated auditory stimuli. Our results demonstrate that attentional loads did not reliably alter the amount of the auditory enhancement effects caused by semantically congruent AV stimuli on this task. However, attentional loads disrupt the auditory enhancement effects of the semantically unrelated and incongruent AV stimuli. These findings highlight a critical role for the semantic association among AV stimuli in modulating the effects of attentional loads on the AV integration of modality-specific selective attention.

3.1 Background

Audio-visual (AV) integration is the phenomenon by which stimuli from visual and auditory sensory modalities can be integrated into a coherent representation to better perceive information [56]. Importantly, many studies have shown that attention plays an important role in the AV integration processing [5,18,20]. It has been demonstrated that it is necessary for attending to at least one unisensory components of a multisensory stimulus for those unisensory stimuli to be integrated, this AV integration cannot occur when subjects were attending away from the multisensory objects [57]. And even though some studies have consistently shown decreases in AV integration when participants are asked to attend selectively to a single modality, relative to distributed attention across both auditory and visual domains [36,57], a number of intersensory attention studies where attention is directed toward one of two sensory modalities, have shown a facilitation of sensory processing for its features in the ignored sensory modality [34,57,58].

When attention is only directed toward one of two sensory modalities, some studies [57,58] in which simple and unrelated auditory and visual stimuli were presented, have shown that attention can automatically spread from an attended visual stimulus to a task-irrelevant, simultaneously presented, auditory stimulus. In the present study we were interested more specifically in how visual selective attention operates on complex naturalistic common objects with well-known multisensory attributes. More recently, a number of intersensory attention studies in which a central stream of alternating pictures and sounds of common objects with apparent higher-level semantic congruence or incongruence were presented, and have shown that attended selectively to a single modality (visual) results in automatic coactivation of that object's representations in ignored sensory modalities [34,35]. Furthermore, Fiebelkorn et al.

(2010) have reported that the degree of semantic association between the relevant visual input and the irrelevant auditory one modulates the cross-modal interaction of audiovisual stimuli of common objects when attention is directed toward visual modality [59]. Results indicated an increased auditory ERP negativity beginning around 200 ms when the response to the ignored sound semantically matched with the attended visual object (e.g. barking sound and dog picture) relative to when the sound that semantically mismatched with the attended visual object (e.g. barking sound and guitar picture). That is, the object-based spread of visual attention to the unattended auditory modality was greater when the visual and auditory semantic features are semantically matched relative to mismatched [59]. Thus, the degree of semantic association between the unimodal components of a multisensory signal may be a factor that influences the extent of top-down attentional controls on audiovisual integration of common objects.

In fact, the influence of attention on the integration of input from different sensory modalities is complex and multifaceted [18-19], and whether the occurrence of multisensory integration is relatively automatic and not affected by increased attentional loads has become a crucial question [20-21]. Even though the fact that the AV integration of common objects was observed even when auditory modality was instructed to ignore may indicate that this integration takes place automatically, regardless of attentional control, it is possible that spare attentional resources may have contributed to this AV integration because the amount of attentional resources have not been effectively controlled. Moreover, prior research revealed that the ignored distractor processing depends critically on the level and type of load involved in the processing of goal-relevant information [22]. Load Theory posited that irrelevant auditory stimuli may be particularly hard to ignore under low load conditions due to spare capacity remaining after processing relevant information spills over to allow processing of other irrelevant distractors; however, irrelevant auditory stimuli can be successfully ignored under higher load conditions, in which the relevant processing exhausts attentional resources [22-23]. Thus, if the AV integration processing under modality-specific selective attention can also occur under high load, the ignored task-irrelevant distractors would presumably have the ‘special’ quality that automatically influences the target detection. Indeed, a number of studies have employed the dual task design in which a distracter task is adopted to modulate the levels of endogenous attentional resources available for the secondary task to directly explore the effect of attentional load on the speech AV integration [28-30] and the integration of emotional AV information [27]. Furthermore, it has been investigated how increased attentional loads influence the AV integration of common objects when attention was distributed across modalities (auditory and visual) [60]. However, it is not clear how the integration of AV stimuli of common objects is influenced by increased attentional loads when only attended selectively to visual modality. Moreover, another crucial question that arises is whether semantic association among AV stimuli modulates the effect of increased attentional loads on the AV integration of common objects when only attended selectively to visual modality.

Thus, the aim of the present study was to explore whether the cross-modal interaction of AV stimuli can occur automatically, is not restricted by increased attentional loads when attended selectively to visual modality, and whether semantic association among AV stimuli modulates the effect of increased attentional loads on the AV integration of common objects in the setting of attended selectively to visual modality. In the present study, we used a dual task paradigm to resolve these questions. And we will adopt an RSVP stream as the distractor task to manipulate different attentional loads: no attentional load, low attentional load, and high attentional load. The participant was presented with an RSVP stream and either asked to ignore it (no-load), detect infrequent yellow letters (low-load), or detect infrequent white numbers (high-load). Similar RSVP streams have been utilized to construct different level of attentional loads in previously published dual task studies [13,14,30]. In the AV interaction task, participants were instructed to respond to a specific image of a common object (e.g., dogs) while ignoring all sounds (i.e., the sounds of animals and white noise). And we used a small stimulus set consisting of semantically congruent multisensory objects (e.g., barking dogs) and semantically incongruent multisensory objects (e.g., birds consistently paired with barks), as well as semantically unrelated multisensory objects (e.g., the images of dogs paired with white noise) to construct different degree of semantic association between the unimodal components of a multisensory signal.

3.2 Methods

3.2.1 Participants

We recruited a total of 20 (seven females, mean age of 25) participants to conduct this experiment. All participants were able to perform AV integration tasks under all load conditions, and the accuracy of both integration processing and distractor tasks was higher than 70% under all load conditions. Participants reported normal to corrected-to-normal hearing and vision. All participants provided written informed consent, and the experiment procedures were previously approved by the ethics committee of Okayama University.

3.2.2 Apparatus and materials

We use the MATLAB software (R2014b, MathWorks, MA, Psychtoolbox, 3) to display the experimental stimuli and record the participants' responses. The visual stimuli were presented on a black background on a 24-inch VG 248LCD (made by ASUS, Taiwan) computer monitor (screen resolution of 1920×1080, refresh rate of 120 Hz) located in a dark and sound attenuated room. The distance between the computer monitor and the participant's head was approximately 57 cm. Auditory stimuli were presented through speakers located on the central monitor. Additionally, two speakers (Harman/Kardon HK206, frequency response: 90-20,000 Hz) were used to present the auditory stimuli. The visual stimuli in the AV integration task consisted of the presentation of animal pictures (17.5 × 12.5 cm), and they are located directly above the center of the computer monitor (located 6 cm from the center of the screen, at a visual angle of approximately 6°) (see Figure 2). There were line drawings of five animal pictures in our experiment: a dog, a bee, a frog, a bird, and a pig. They were taken from Snodgrass and Vanderwart (1980) and were standardized on familiarity and complexity [37]. The sounds of these five animals were collected through internet searches and later standardized and modified such that each single animal sound had a duration of 300ms. The animal sounds and white noise were presented at a comfortable listening level of ∼75 dB SPL.

Furthermore, the pictures and sounds of animals were combined to form semantically congruent AV pairs (e.g., "dog" paired with "bark"), semantically incongruent AV pairs (e.g., "dog" paired with "tweet") and semantically unrelated AV pairs (e.g., "dog" paired with "white noise") (see Figure 1). Finally, four stimulus types were included in this study, namely, animal pictures alone(V stimuli) ; a combination of pictures and sounds belonging to the same animal (Congruent AV stimuli); a combination of pictures and sounds belonging to different animals (Incongruent AV stimuli); and a combination of animal pictures and white noise (Unrelated AV stimuli). The target stimuli were the pictures of "dog" or the AV stimulus including "dog" pictures regardless of the accompanying sound (see Figure 9).

The visual distractor set in the Rapid Serial Visual Presentations (RSVP) task consisted of 23 letters (A, C, D, E, F, J, H, J, K, L, M, N, P, Q, R, S, T, U, V, W, X, Y, Z) and seven digits (2, 3, 4, 5, 6, 7, 9), they are 2 × 2 cm, subtending a visual angle of 3.3° × 1.9° (see Figure 10). The remaining digits and letters were not adopted because some letters (I, B, O) and digits (1, 8, 0) were so similar that they may have an unbalanced impact to participants relative to other distractor stimuli [13,14,30].

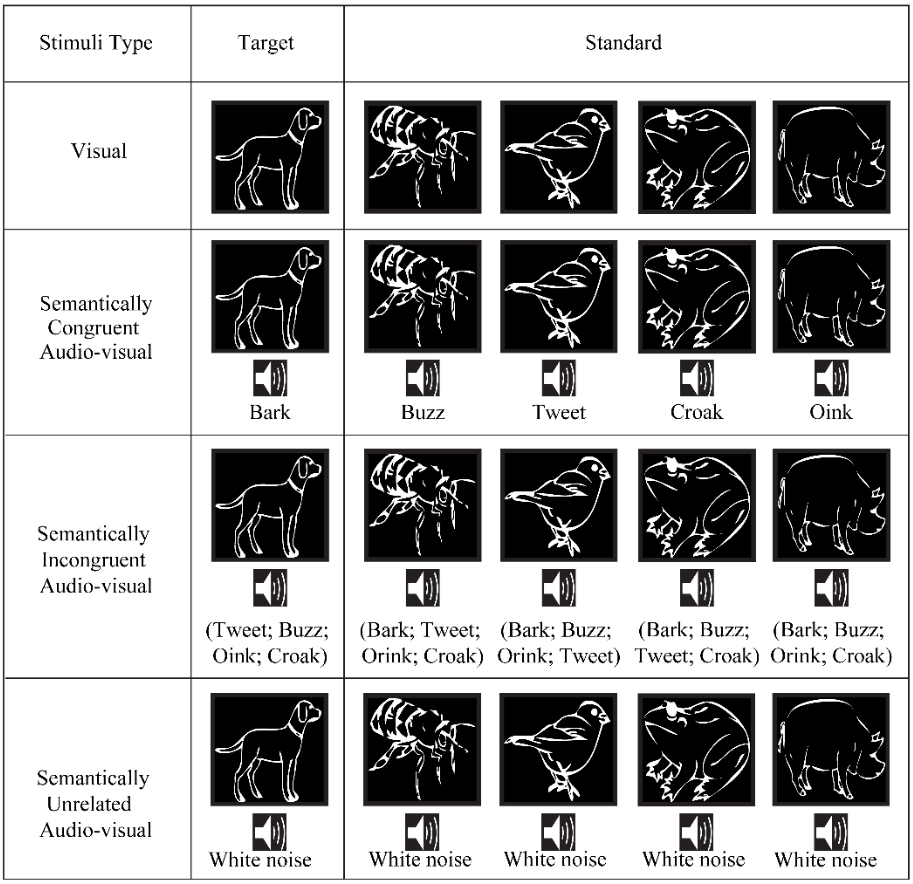

Fig. 9 *The stimulus type used in the animal identification tasks. In this study, four stimulus types were included: animal pictures alone; a combination of pictures and sounds belonging to the same animal; a combination of pictures and sounds belonging to different animals; a combination of pictures and white noise.*

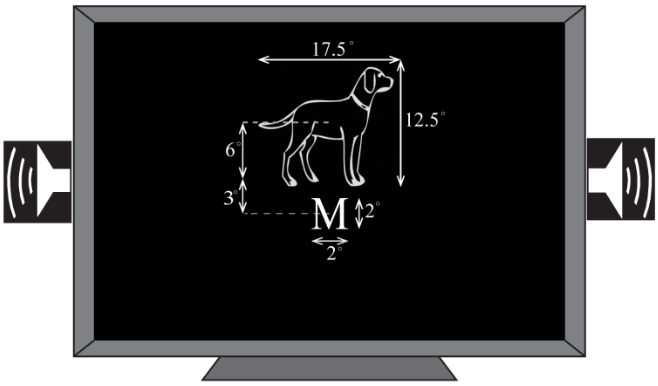

Fig. 10 *The orientation of the visual stimuli of the animal identification task and the distractor task (RSVP task) which are presented simultaneously during the experiment.*

3.2.3 Procedure

Participants were seated in a comfortable chair in a dimly lit, electrically shielded, and sound-attenuated room (laboratory room, Okayama University, Japan) with their head positioned on a chin rest to perform a dual task. The factorial design had two within subject factors: Stimuli modality (V, Congruent AV, Incongruent AV, Unrelated AV), and Attentional load (No-load, Low-load, High-load). First, a small stimulus set consisting of semantically congruent multisensory objects (e.g., barking dogs) and semantically incongruent multisensory objects (e.g., birds paired with barks), as well as semantically unrelated multisensory objects (e.g., the images of dogs paired with white noise) were used to construct different degree of semantic association between the unimodal components of a multisensory signal.

Second, participants were only asked to judge the visual targets of the animal identification task (AV integration task, ignore all the auditory stimuli) and not to perform the distractor task (no load), or simultaneously perform the animal identification task with a distractor task that requires participants to search a central RSVP stream for either a yellow letter (low load) or a white digit (high load) [13,14,30]. Importantly, the task of searching for digits in a series of letters under high-load condition is a higher level of semantic processing, while the task of searching for yellow letters under low-load condition is the judgment of physical property of an object. Therefore, the digits searching task under the high-load condition is more difficult and requires more attentional resources than the task of searching only for a specific colour under the low-load condition. In this way, by increasing the difficulty of distractor task, we can control the attentional resource that is used during AV integration processing. Hence, our study included three load condition types, namely, no load, low load, and high load.

In the high-load condition, the RSVP white digits detection task and the animal identification task (AV integration task) are displayed simultaneously [13,14,30] (see Figure 11). For the animal identification task, one animal picture (300ms) is randomly presented in each trial, and each trial began with a 400ms presentation of the fixation cross to indicate the beginning of a trial. The animal pictures were consistently presented directly above the RSVP streams, and the pictures and sounds of animals were randomly presented in the position of the first through fifth letter of RSVP stream for a period of 300ms, and the stimulus onset asynchrony (SOA) between the onset of fixation at the beginning of each trial and animal pictures was 400-1450 ms. Visual-alone, congruent AV, incongruent AV stimuli and unrelated AV stimuli were presented equiprobably and in pseudorandom order to limit predictability. If the animal image presented is a “dog” picture, either appearing alone or co-occurring with an auditory stimulus, subjects should press the “F” key on the keyboard as soon as possible. At the end of a total trial, a blank interface (1000ms) was presented to ensure that the subject had sufficient time to respond to the animal identification task. A total of 640 trials (320 target trials and 320 task-irrelevant trials) were included under each attentional load condition in the experiment. To avoid the fatigue, these trials were divided into 4 main blocks of 160 trials each under each load condition. Each block contained 40 unimodal visual stimuli, 120 audiovisual stimuli (Congruent AV, Incongruent AV, Unrelated AV), and lasted approximately 10 min. The task-irrelevant stimuli constituted 50% of the total stimuli.

For the RSVP white digits detection task, each trial consisted of the presentation of a stream of 7 alphanumeric characters which were continuously displayed at a rate of 8 Hz. Specifically, these different letters were sequentially presented, being randomly replaced every 150 ms. We randomly select the distractor letters in the stream before each trial, but there are no repeaters in the given stream. The target of the RSVP task was presented equiprobably in the first through seventh positions in the stream. The letters in the stream were chosen randomly prior to each trial, with the sole restriction being that no distractor was repeated within a given stream. Additionally, in each trial, either a yellow letter target, a dog target, both, or never both was presented. Specifically, the RSVP streams in each trial had a 25% probability of containing no numbers or yellow letters, a yellow letter only, a number only, or a yellow letter and a number, thus resulting in a 50% probability of a target being present in each trial for all attentional load conditions. With respect to the RSVP task, participants were asked to respond at the end of each trial, that is, after the interface of the red gaze point appeared, the subjects were to press the “J” button within 1000 milliseconds if they observed a target during the RSVP task.

In the low-load condition, except for the target stimulus of RSVP task is yellow letters, the other experimental conditions are the same as the high load [13,14,30]. They were instructed not to respond to any digit target that might appear in the RSVP stream.

In the no-load condition, although the rapid serial visual presentation (RSVP) task was also displayed simultaneously with the animal identification task, participants were asked to ignore of the RSVP task [35]. Participants did not need to judge the RSVP task, only to determine whether the animal image presented was a “dog”.

The stimulus sequence presented is the same for all load conditions, each load condition was completed in four separate blocks, and participants were permitted to take breaks between blocks. The order that participants completed the attentional load condition blocks was randomized and counterbalanced across participants. Before the experiment was officially started, all participants engaged in a practice experiment with 16 trials to ensure that they correctly understood the experimental procedures and responded correctly to the different tasks. The entire experimental session lasted about 120 min.

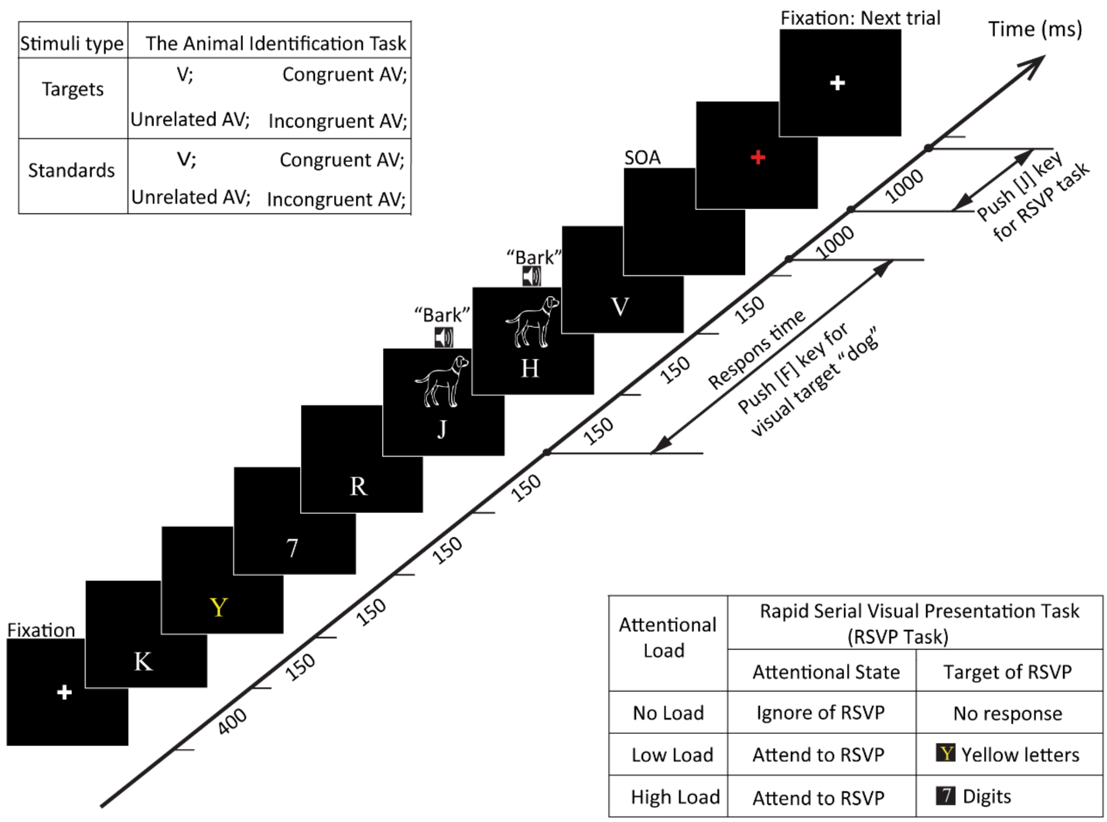

Fig. 11 *A schematic representation trial in which both the animal identification task and the RSVP task are run simultaneously. Participants must judge whether the animal image presented is a "dog" while viewing a stream of letters, either ignoring the RSVP streams (no-load), reporting yellow letters (low-load), or reporting numbers (high-load). In each trial, a 400ms gaze point is presented to inform the beginning of the new trial. Seven letters (may contain a number) are then continuously displayed at a rate of 8 Hz. The picture or sound of an animal is randomly presented in the position of the first to fifth letter of the RSVP stream. Their presentation time is 300ms. With respect to the animal identification task (ignore auditory stimuli), participants should respond as soon as possible to the "dog" pictures by pressing the "F" key on the computer keyboard. After the presentation of seven letters, a blank interface (1000ms) is presented to ensure that the subject has sufficient time to react to the animal identification task. Regarding the RSVP task, if a target for this trial is observed, the participant should press the "J" key on the computer keyboard when the interface of the red gaze point appears (1000ms).*

3.2.5 Data analysis

(1) Response time and accuracy analysis of the AV integration task

Participants' median accuracies and response times (RT) were calculated for the stimuli of different modalities under different attentional load condition after removing RT outliers (1.8%). Incorrect trials and trials with response times shorter than 200 ms or longer than 1100 ms were removed. To determine whether the cross-modal interaction of AV stimuli which have different degree of semantic association between the unimodal components were successfully elicited and whether increased attentional load influenced these cross-modal interaction, we conducted a repeated-measures analysis of variance with median response times as the dependent factor and Stimulus modality (Visual, Congruent AV, Incongruent AV, Unrelated AV) and Attentional load (No-load, Low-load, High-load) as independent factors. If the Shapiro-Wilk test was significant, percentage of accuracy was analyzed by nonparametric repeated-measures analyses of variance (ANOVA; the Friedman test). $P < 0.05$ was considered to be statistically significant.

(2) The analysis of auditory enhancement effect in the AV integration task

In addition, the auditory enhancement effects represent the percent change in performance for AV stimuli compared to visual stimuli [61], and the amount of improvement in the response times due to the addition of auditory stimuli were calculated by formula as follows:

$$\text{Auditory enhancement effect (\%)} = (RT_V - RT_{AV})/RT_V \times 100\% \quad (2)$$

RT_{AV} is the response times of the correctly detected AV stimulus and RT_V is the response times of the correctly detected visual stimulus. Furthermore, to test whether the auditory enhancement effects produced by semantically congruent AV stimuli were influenced by attentional load, we conducted a repeated-measures analysis of variance on the auditory enhancement effects under different attentional load conditions. We use the same method to test whether attentional loads have a different effect on the auditory enhancement effects of semantically unrelated AV stimuli or semantically incongruent AV stimuli. In addition, the degrees of freedom were corrected using the Greenhouse-Geisser correction when necessary.

(3) Calculation of Cumulative Distribution Functions of AV integration task

To assess whether the magnitudes of the auditory enhancement effects were reduced by attentional loads and whether there was a difference under each load condition, we performed further analysis as follows. First, we subtracted the CDFs for responses to unisensory visual targets from the CDFs for responses to matching multisensory targets to obtain a measure of the auditory enhancement effects of semantically congruent AV stimuli [62]. We compared the observed RTs of the congruent AV CDF of each participant in each load condition to the visual CDF at each time bin to test for probability difference by performing planned pairwise comparisons for different attentional load conditions (No-load, Low-load, or High-load). Probability difference between congruent AV CDF and V CDF (i.e., $RT_{\text{Congruent AV}} < RT_{V}$) indicated the auditory enhancement that exceed statistical facilitation. Two-tailed P values were corrected for multiple comparisons using the Bonferroni correction, and differences are reported as significant at $p < 0.05$.

In addition, to test for differences in the positive area under the difference curve (i.e., the difference in probability of the semantically congruent AV CDF and the visual CDF for the RT range of 200 to 1100 ms) between attentional load, the positive area under the difference curve across the 10-ms time bins of the CDF of each load type of each participant was used in a repeated-measures ANOVA with the factor Attentional load (No-load, Low-load, High-load), followed by planned pairwise comparisons. In order to extract the positive area under the curve for each participant, the difference curve between the congruent AV CDF and the visual CDF was calculated for each participant. Next, all negative probabilities (no violation between congruent AV and V) were set to a value of zero and only the positive area under the curve was calculated for all participants. Furthermore, the greatest AV facilitation is defined as peak benefit, and the time spanning from the presentation of the target to the peak benefit is defined as the peak latency (see Fig. 14).

We subtracted the CDFs for responses to unisensory visual targets from the CDFs for responses to semantically incongruent AV targets to obtain a measure of the auditory enhancement effects produced by incongruent AV stimuli (Laurienti et al., 2004; Mozolic et al., 2008). We adopted the same method as above to test for the positive area under the curve (i.e., the difference in probability of the semantically incongruent AV CDF and the visual CDF for the RT range of 200 to 1100 ms) between different attentional loads. In addition, we adopted the same method to explore whether increased attentional loads would influence the enhancement effects of semantically unrelated AV stimuli. The CDFs for responses to unisensory targets were subtracted from the CDFs for responses to unrelated AV targets to obtain a measure of the enhancement effects of semantically unrelated AV stimuli.

(4) Impact analysis of the distractor task

First, the RSVP performance is checked to verify that participants accurately performed the distractor task (because they could have simply ignored it and only attended the primary task). A Shapiro-Wilk test was conducted to confirm the assumption of a normal distribution in low-load and high-load conditions. If the Shapiro-Wilk test was not significant, the repeated-measures ANOVA for comparisons between different load conditions were conducted. If the Shapiro-Wilk test was significant, we used the one-way nonparametric repeated-measures analyses of variance (ANOVA; the Friedman test) for comparisons. $P<0.05$ was considered to be statistically significant. Second, we calculated the relative performance under the no-load, low-load and high- load conditions for all stimuli modality (Visual, Congruent AV, Incongruent AV, Unrelated AV) to explore whether attentional loads significantly disrupted the RTs for the AV integration task. In addition, in each analysis, the degrees of freedom were corrected using the Greenhouse-Geisser correction when the Mauchly's test indicated that the assumption of sphericity had been violated.

3.3 Results

3.3.1 Response Times of the AV integration task

To determine how increased attentional load interacts with semantic congruency to influences AV integration, we conducted repeated-measures ANOVA on median response time using Stimulus modality (Visual, Congruent AV, Incongruent AV, Unrelated AV) and Attentional load (No-load, Low-load, High-load) as factors. The results revealed a significant main effect of stimulus type, $F\ (2.654, 50.433) = 31.562$, $p < 0.001$, $\eta^2 = 0.624$, participants responded significantly more slowly on visual target trials (M = 461 ms, SE= 8) than on semantically congruent AV (M = 440 ms, SE=7; $p < 0.001$) or semantically unrelated AV (M = 453 ms, SE=7; $p = 0.017$) target trials, but not significantly slower than semantically incongruent AV (M = 456 ms, SE=7; $p = 0.063$) target trials. Moreover, the main effect of attentional load condition was also significant [$F\ (1.541, 29.281) = 58.59$, $p <0.001$, $\eta^2=0.755$], with the participants responding more rapidly overall in the no-load condition (M = 399 ms, SE=7) than in the low-load (M = 463 ms, SE= 9) and high-load conditions (M = 496 ms, SE=11) as expected. Interestingly, we also found a significant interaction between modality and attentional load [$F\ (4.656, 88.472) = 5.457$, $p < 0.001$, $\eta^2= 0.233$]. Post hoc subsidiary analyses with Bonferroni adjustment for multiple comparisons demonstrated that the median response times for the semantically congruent AV target trials were significantly faster than were the visual response times under each load condition [no-load: $t_{(19)} = 8.349$, $p = 0.002$; low-load: $t_{(19)} = 7.847$, $p = 0.003$; high-load: $t_{(19)} = 5.53$, $p = 0.004$] (see Figure 12). However, the median response times for unrelated AV trials were significantly faster than those for visual trials under the no-load and low-load conditions [no-load: $t_{(19)} = 5.725$, $p < 0.001$; low-load: $t_{(19)} = 2.889$, $p = 0.056$], but has no significant difference between unrelated AV trials and visual trials under high-load condition [high-load: $t_{(19)} = 0.477$, $p = 0.639$] (see Figure 4C). Additional, the median response times for incongruent AV trials were significantly faster than those for visual trials under the no-load, low-load conditions [no-load: $t_{(19)} = 3.671$, $p = 0.01$; low-load: $t_{(19)} = 3.977$, $p = 0.005$], but it has no difference between visual and incongruent AV trials under high-load condition [high-load: $t_{(19)} = -0.599$, $p = 0.556$] (see Figure 12B).

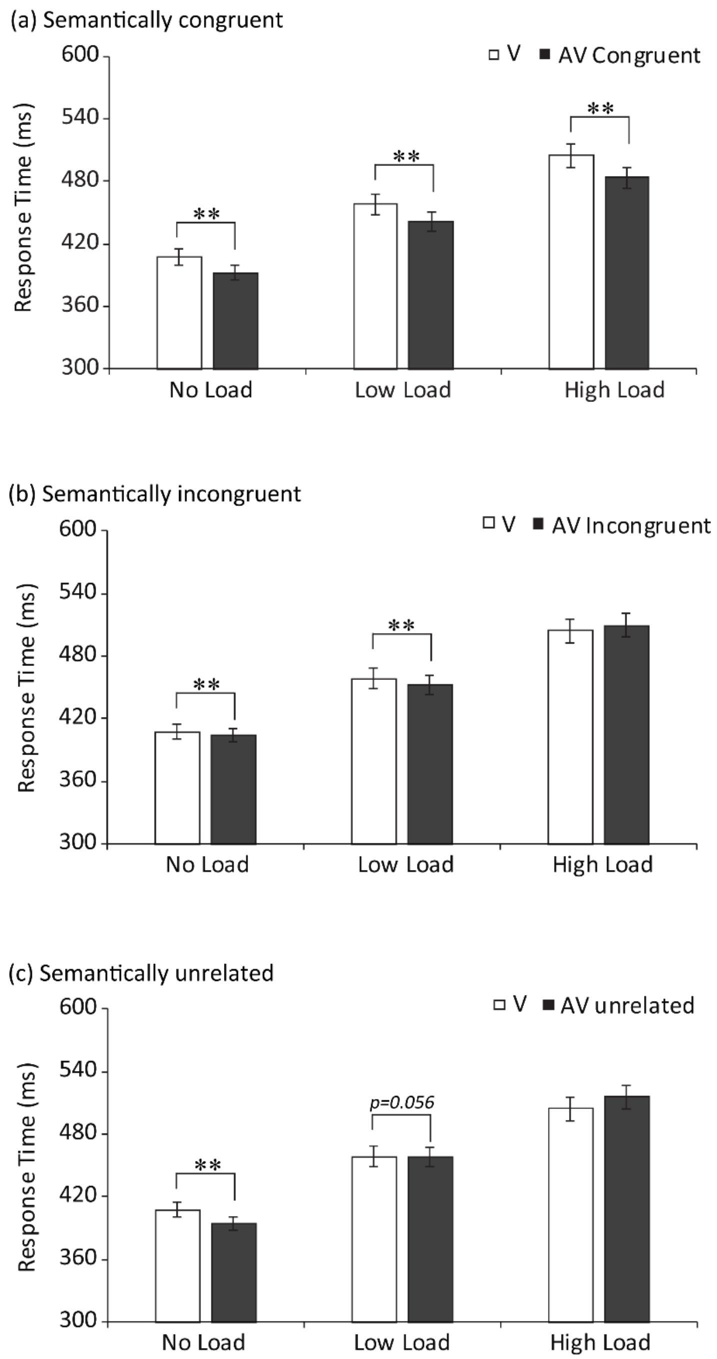

Fig. 12 *Median response times for the visual and semantically related AV trials (a); visual and semantic unrelated AV trials (b); and visual and semantic incongruent AV trials (c) in the animal identification task under different attentional load conditions. Error bars represent the standard errors of the means. ***p < 0.001, ** p < 0.01, * p < 0.05.*

3.3.2 Auditory enhancement effect in the AV integration task

Furthermore, to test whether the auditory enhancement effects produced by semantically congruent AV stimuli were influenced by attentional loads, a repeated measures ANOVA was used. The repeated-measures ANOVA did not reveal a main effect of attentional load [$F (1.946, 36.978) = 0.413$, $p = 0.659$, $\eta^2=0.021$]. Post hoc paired t tests with the Bonferroni correction showed that attentional loads does not significantly disrupt this auditory enhancement effects [no-load/low-load: $t_{(19)} = -0.949$, $p = 0.355$; no-load/high-load: $t_{(19)} = -0.275$, $p = 0.786$; low-load/high-load: $t_{(19)} = 0.610$, $p =0.549$] (see Figure 12).

To test whether the auditory enhancement effects of semantically unrelated AV stimuli or semantically incongruent AV stimuli were influenced by attentional loads, the repeated-measures ANOVA were conducted. For the auditory enhancement effects of semantically incongruent AV stimuli, the repeated-measures ANOVA revealed a main effect of attentional loads [$F (1.754, 33.323) = 6.515$, $p = 0.004$, $\eta^2=0.255$]. Post hoc paired t tests with the Bonferroni correction showed that the auditory enhancement of incongruent AV trials was significantly larger in the high-load condition compared to the no-load condition [$t_{(19)} = 2.884$, $p = 0.028$] and low-load [$t_{(19)} = 2.870$, $p =0.029$], but there is no significant difference between no-load and low-load condition [$t_{(19)} = 0.434$, $p = 0.669$] (see Figure 13).

For the auditory enhancement effects of semantically unrelated AV stimuli, the repeated-measures ANOVA also revealed a main effect of attentional loads [F (1.731, 32.891) = 8.545, p = 0.001, η^2=0.310]. Post hoc paired t tests with the Bonferroni correction showed that the auditory enhancement of unrelated AV trials was significantly larger in the no-load condition compared to the low-load condition [$t_{(19)}$ = 2.603, p = 0.052] and high-load condition [$t_{(19)}$ = 3.497, p = 0.007], but there is no significant difference between low-load and high-load condition [$t_{(19)}$ = 2.013, p =0.176] (see Figure 13).

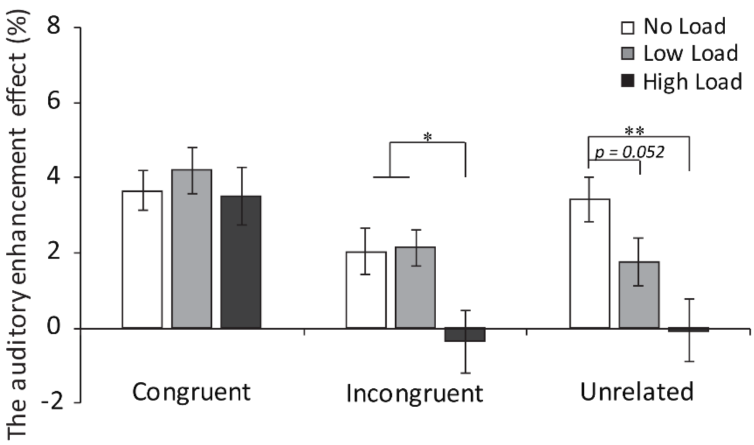

Fig. 13 *Comparison of the magnitudes of the median auditory enhancement effects (%) elicited by semantically congruent AV trials; by semantically incongruent AV trials, and by semantically unrelated AV trials under each load condition. Positive values indicate the improvement, whereas negative values indicate the impairment. ***p < 0.001, **p < 0.01, *p < 0.05.*

3.3.3 Accuracy of the AV integration task

The accuracy in the AV integration task in all load conditions violated the Shapiro-Wilk tests (all $W < 1$, all $p < 0.01$), the non-parametric Friedman tests on the accuracy of AV integration task showed significant differences under different load conditions ($\chi^2(11) = 38.73$, $p < 0.001$). The Wilcoxon signed-rank tests on the coefficient of variance showed no significant influences for some stimulus types in the no-load condition (V vs Unrelated AV, $W_{(20)} = -0.79$, $p = 0.428$; V vs Congruent AV, $W_{(20)} = -0.054$, $p = 0.957$; V vs Incongruent AV, $W_{(20)} = -0.535$, $p = 0.593$), low-load condition (V vs Unrelated AV, $W_{(20)} = -1.303$, $p = 0.193$; V vs Congruent AV, $W_{(20)} = -1.404$, $p = 0.160$) and high-load condition (V vs Unrelated AV, $W_{(20)} = -1.069$, $p = 0.285$; V vs Congruent AV, $W_{(20)} = -0.331$, $p = 0.741$; V vs Incongruent AV, $W_{(20)} = -0.655$, $p = 0.513$). However, there was a significant difference between visual and incongruent AV stimuli in the low-load condition (V vs Incongruent AV, $W_{(20)} = -1.965$, $p = 0.049$). Based on the fact that the RT effects were not due to a speed accuracy trade-off and based on the generally very low error rates, hence, the auditory enhancement of different AV stimuli might be reflected mostly in RTs.

Table 4 *Median accuracy (%) and response times (RTs, ms) with standard deviations (SDs) for the target of each trial type under no -load, low-load, and high-load.*

Stimuli	**No Load**		**Low Load**		**High Load**	
Type	**RTs (SD)**	**Accuracy (SD)**	**RTs (SD)**	**Accuracy (SD)**	**RTs (SD)**	**Accuracy (SD)**
V	407.5(33.2)	100(3.3)	458.2(44.0)	100(2.06)	504.4(50.0)	97.5(3.49)
Congruent AV	392.5(33.3)	98.8(2.84)	441.4(42.8)	98.8(1.74)	483.4(45.9)	98.8(2.16)
Incongruent AV	403.7(29.5)	100(3.64)	451.8(41.3)	98.8(1.43)	509.6(49.1)	98.1(2.90)
Unrelate AV	394.3(28.8)	100(3.05)	458.4(40.0)	98.8(2.70)	515.4(48.6)	98.8(2.12)

3.3.4 Cumulative distribution functions of the AV integration task

To assess the effects of matching cross-modal promoters, we compared the unisensory visual CDF to the matching AV CDF under different attentional load conditions (Figure 14). A comparison between the unisensory visual targets CDFs and semantically congruent AV targets CDFs at each time bin revealed that the auditory enhancement effects were observed under all the load conditions ($p < .05$, paired t-tests, 2-tailed, Bonferroni corrected, Figure 6a). Specially, the auditory enhancement effects occurred was 170 ms long in the no-load condition (310-480 ms). In the low-load condition, this time-window was 230 ms long

(340-570 ms), which was smaller compared to the high-load condition in which this window was 290 ms long (350-640 ms) (Table 2).

Moreover, the positive area under the curve (The subtraction of visual distribution from the congruent AV distribution) was compared between the different load conditions. The repeated-measure ANOVA revealed a main effect of attentional load condition [$F(1.559, 29.618) = 1.409$, $p = 0.257$, $\eta^2 = 0.069$]. Post hoc comparisons (Bonferroni corrected) showed that the positive area under the curve was not significantly larger in the no-load condition (*M* =12.38 ms, *SE* = 1.57) compared to the low-load condition (*M*=9.47 ms, *SE* =1.29, $t_{(19)} = 0.563$; $t_{(19)}=0.56$, $p = 0.58$) and high-load condition (*M*=14.54 ms, *SE* =1.75, $t_{(19)} = -0.904$, $p = 0.377$), and there is also no significant difference between low-load and high-load condition ($t_{(19)} = -2.163$, $p = 0.131$).

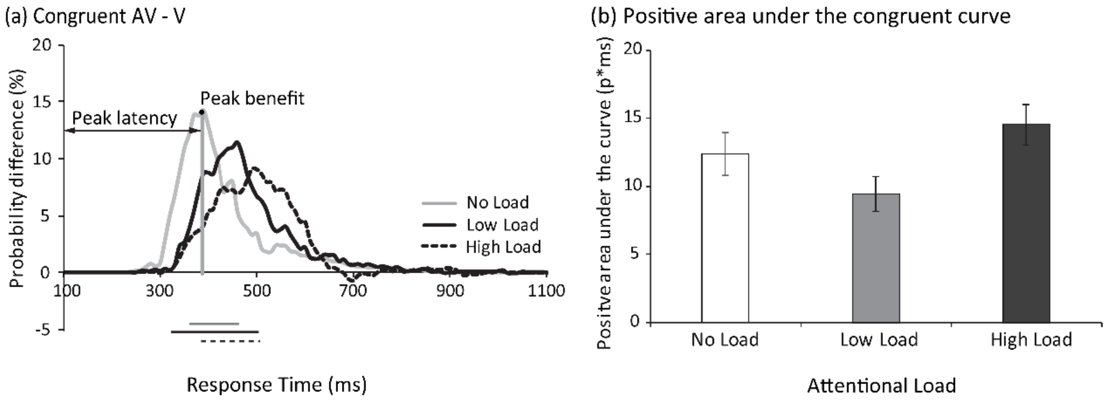

Fig. 14 *(a) The subtraction of the visual distribution from the congruent AV targets distribution across the full range of response times (RTs) for no load (solid grey line), low load (black dotted curve), and high load (black dashed curve). Significant violations are indicated with a horizontal bar in the graph below the x-axis indicating the RT range in which the probabilities of congruent AV targets of consecutive time points are larger than the probabilities of visual targets. This result shows that significant auditory enhancement is present under all load conditions. (b) Median positive area under the curve (The subtraction of visual distribution from the congruent AV distribution) in each of load conditions. Significant differences are indicated with an asterisk (p < 0.05).*

To assess the effects of non-matching cross-modal distractors, we compared the response distributions for unisensory visual trials to the response distributions for non-matching AV trials under different perceptual conditions (Figure 15). The comparison between the unisensory visual targets CDF and semantically incongruent AV CDF in each time bin showed an auditory interference effect, however, this auditory interference effect was only observed under no-load condition and low-load condition ($p < 0.05$, paired t-tests, 2-tailed, Bonferroni corrected, Figure 15a). Specially, an auditory enhancement effect was observed at

310-450 ms in the no-load condition ($p < 0.05$); and a significant auditory enhancement effect also be found at 360-500 ms in the low-load condition ($p < 0.05$), but no auditory interference effect was found in the high-load condition. The negative area under the curve was also compared between the different load conditions (see Fig. 15, right panel).

The repeated-measures ANOVA revealed a main effect of load condition [$F\ (1.398, 26.569) = 8.432$, $p = 0.004$, $\eta^2 = 0.307$]. Post hoc comparisons (Bonferroni corrected) showed that the positive area under the curve was not significantly larger in the no-load condition (*M* =6.41 ms, *SE* =1.69) compared to the low-load condition (*M*=7.23 ms, *SE* =0.97, $t_{(19)} = 0.195$, $p = 0.847$), but the no-load condition was significant larger than high-load condition (*M*=2.27 ms, *SE* =0.30, $t_{(19)} = 3.496$, $p = 0.007$), besides, the low-load condition was also significantly larger than high-load condition ($t_{(19)} = 5.51$, $p < 0.001$).

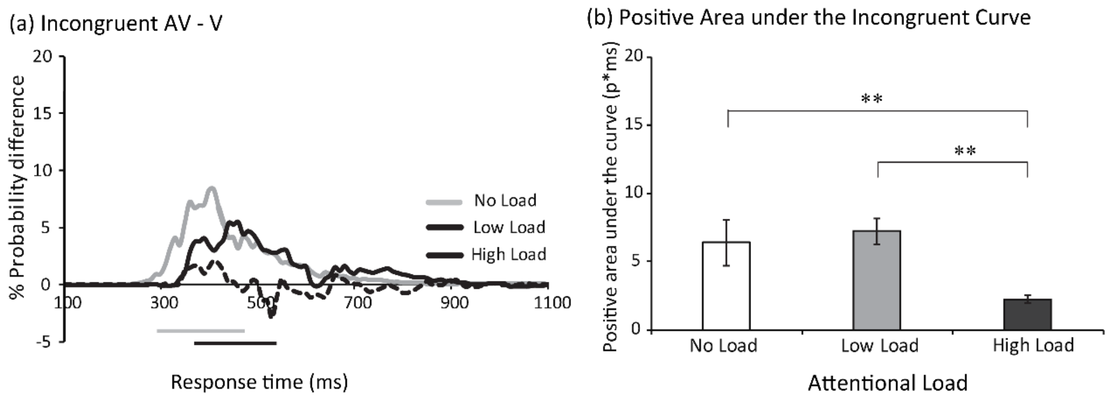

Fig. 15 *(a) The subtraction of the visual distribution from the incongruent AV targets distribution across the full range of response times (RTs) for no load (solid grey line), low load (black dotted curve), and high load (black dashed curve). Significant violations are indicated with a horizontal bar in the graph below the x-axis indicating the RT range in which the probabilities of incongruent AV targets of consecutive time points are larger than the probabilities of visual targets. This result shows that significant auditory enhancement is only present under no load and low load conditions. (b) Median positive area under the curve (The subtraction of visual distribution from the incongruent AV distribution) in each of load conditions. Significant differences are indicated with an asterisk (p < 0.05).*

To assess the effects of unrelated cross-modal promoters, we compared the unisensory CDF to the unrelated multisensory CDF under different attentional load conditions (Figure 16). A comparison between the unisensory visual targets CDFs and semantically unrelated AV targets CDFs in each time bin revealed the auditory enhancement effects under the no-load, low-load and high-load conditions ($p < 0.05$, paired t-tests, 2-tailed, Bonferroni corrected, Figure 8a). Specially, the auditory enhancement effects occurred from 320 ms to 450 ms in the no-load condition ($p < 0.05$), and from 550 ms to 700 ms in the low-load condition ($p < 0.05$), but no auditory enhancement effect was found in the high-load condition (Table 2).

The positive area under the curve between unrelated AV CDF and V CDF was also compared between the different load conditions (see Fig. 16, right panel). The repeated-measures ANOVA did not reveal a main effect of load condition [F (1.604, 30.47) =4.74, $p = 0.022$, $\eta^2 = 0.200$]. Post hoc comparisons (Bonferroni corrected) showed that the positive area under the curve was significantly larger in the no-load condition (*M* =9.11 ms, *SE*=1.43) compared to high-load condition (*M=3.93 ms, SE =0.90*, $t_{(19)} = 3.66$, $p = 0.005$), but the no-load condition was not significant larger than the low-load condition (*M*=8.00 ms, *SE* = 2.09, $t_{(19)} = -0.115$, $p = 0.909$). In addition, there is a borderline significant difference between low-load and high-load condition ($t_{(19)} = 2.49$, $p = 0.067$).

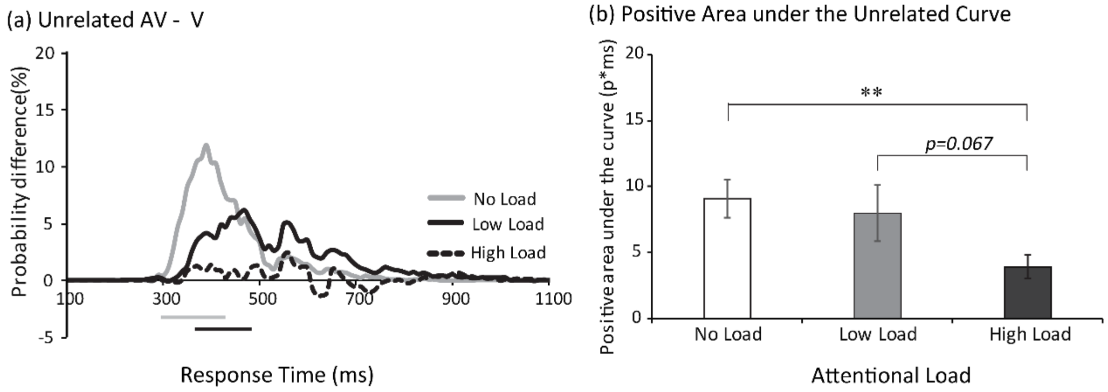

Fig. 16 *(a) The subtraction of the visual distribution from the unrelated AV targets distribution across the full range of response times (RTs) for no load (solid grey line), low load (black dotted curve), and high load (black dashed curve). Significant violations are indicated with a horizontal bar in the graph below the x-axis indicating the RT range in which the probabilities of unrelated AV targets of consecutive time points are larger than the probabilities of visual targets. This result shows that significant auditory inhibition is only present under no load and low load conditions. (b) Median positive area under the curve (The subtraction of visual distribution from the unrelated AV distribution) in each of load conditions. Significant differences are indicated with an asterisk (p < 0.05).*

3.3.5 Impact of the distractor task (RSVP task)

First, because the Shapiro-Wilk test for the accuracy of the RSVP task under each load condition was not significant (low-load: W = 0.957, p = 0.485; high-load: W = 0.935, p = 0.195), we conducted the repeated-measures ANOVA to determine whether accuracy in the RSVP task was reduced by attentional loads. The results indicated that the accuracy of the RSVP task was significantly higher under the low-load condition (*M* = 94.9 %, *SE* = 0.24) than that under the high-load condition (*M* = 91.4 %, *SE* = 0.45) [$F(1,19) = 88.694$, $p < 0.001$, $\eta^2 = 0.824$]. Moreover, the accuracy of the RSVP performance was above 90%, indicating that the participants accurately performed the distractor task.

Again, the repeated-measures ANOVA using Stimulus modality (Visual, Congruent AV, Incongruent AV, Unrelated AV) and Attentional load (No-load, Low-load, High-load) as factors in the AV integration task revealed a main effect of load [F (1.541, 29.281) = 58.59, p <0.001, η^2=0.755], the post-hoc test showed that inter-participant median RTs for the AV integration task were significantly slower under the low-load (*M* = 463, *SE* = 9) compared with the no-load (*M* = 399, *SE* = 7, $t_{(19)}$= 8.0, $p < 0.001$) condition, and the median RTs under the high-load (*M* = 496, *SE* = 11) were slower than those under the low-load ($t_{(19)}$= 4.7, $p = 0.001$) condition. Furthermore, it is further determined that attentional load significantly disrupts the response times regarding visual stimuli [no-load/low-load: $t_{(19)} = -7.11$, $p <0.001$; no-load/high-load: $t_{(19)} = -7.07$, $p < 0.001$; low-load/high-load: $t_{(19)} = -4.0$, $p = 0.003$] as well as congruent AV stimuli [no-load/low-load: $t_{(19)} = -6.44$, $p <0.001$; no-load/high-load: $t_{(19)} = -7.16$, $p <0.001$; low-load/high-load: $t_{(19)} = -4.0$, $p = 0.001$], incongruent AV stimuli [no-load/low-load: $t_{(19)} = -8.0$, $p < 0.001$; no-load/high-load: $t_{(19)} = -9.45$, $p < 0.001$; low-load/high-load: $t_{(19)} = -5.0$, $p <0.001$] and unrelated AV stimuli [no-load/low-load: $t_{(19)} = -8.75$, $p <0.001$; no-load/high-load: $t_{(19)} = -9.54$, $p <0.001$; low-load/high-load: $t_{(19)} = -4.375$, $p = 0.002$] (see Figure 17). In summary, the response times to all stimulus modalities were significantly slower under high-load condition than under no-load and low-load conditions (all $F > 1$, all $p < 0.01$). Hence, these results indicated that the high-load condition was more demanding than no-load and low-load conditions.

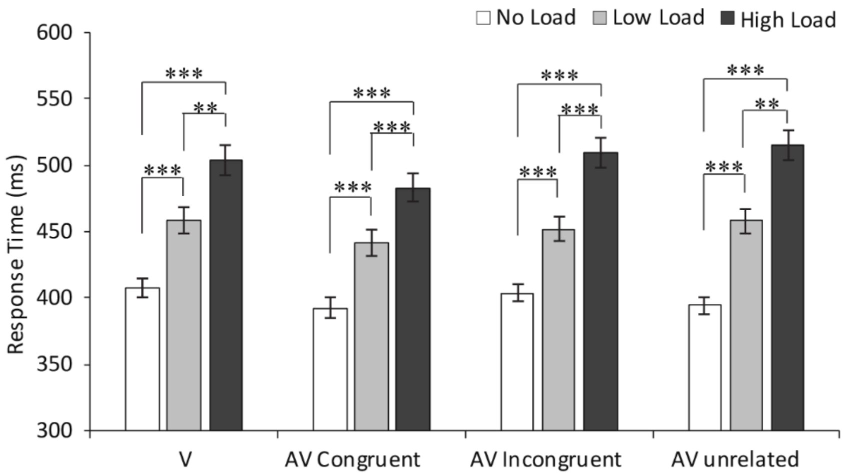

Fig. 17 *The median RTs under the unimodal A, bimodal congruent AV, bimodal incongruent AV and bimodal unrelated AV conditions are presented under different load conditions. The response times to all stimuli generally increased as the load increased. The error bars represent the standard error of the mean. ***p < 0.001, **p < 0.01, *p < 0.05.*

Overall, these results demonstrated two key findings. First, participants accurately performed the task under all load conditions, and the load manipulation was indeed functional. Second, the load manipulation indeed interfered with the target processing in the AV integration task because the response times to all target stimuli was significantly decreased by attentional load.

3.4. Discussion

In the current study, we have investigated how the cross-modal interaction processing of semantically congruent, incongruent and unrelated AV stimuli of complex naturalistic common objects is influenced by different levels of attentional loads in the setting of attended selectively to visual modality. We used an RSVP task to manipulate the amount of attentional resources that were available for the processing of cross-modal interaction. Our main result was that increased attentional loads did not influence the auditory enhancements associated with semantically congruent AV stimuli (see Figures 12 , 13 and 14), but eliminated the auditory enhancements associated with semantically incongruent AV stimuli (see Figures 12 , 13 and 15) and semantically unrelated AV stimuli (see Figures 12 , 13 and 16). Our finding provides strong evidence that highly learned associations among the multisensory features of a common object, formed through a lifetime of experiences with that object, modulate the effect of increased attentional loads on the cross-modal interaction of multisensory features of common objects in the setting of attended selectively to visual modality.

An interesting behavioral finding in our study is that all the semantically congruent, incongruent and unrelated AV detection performance were enhanced relative to isolated visual detection when attended selectively to visual modality under no load condition (see Figure 12). Furthermore, increased attentional loads did not influence the auditory enhancements associated with semantically congruent AV stimuli (see Figures 12 , 13 and 14), but eliminated the auditory enhancements associated with semantically incongruent AV stimuli (see Figures 12, 13 and 15) and semantically unrelated AV stimuli (see Figures 12, 13 and 16). Our finding may be explained according to the “attentional load theory”, which postulates that in the attentional-load conditions, the participants were instructed to response to the AV integration task and additional distractors of RSVP task simultaneously, which lead to that the attentional resources were largely occupied by RSVP task, exhibiting significantly less spare attentional resources for the perception of task-irrelevant auditory stimuli than no load condition. Furthermore, the task of searching for digits in a series of letters under high-load condition requires a higher level of semantic processing and more attentional resources than the task of searching only for a specific colour under the low-load condition. Therefore, more attentional resources were diverted to the competing RSVP task under high-load condition, which would leave few spare attentional resources for the perception of task-irrelevant auditory stimuli in the AV discrimination task simultaneously, thus leads to the influence of ignored auditory stimulus on visual target detection is diminished.

For semantically congruent AV trials, based on the reason that the mental representations of different sensory patterns belonging to the same object have been firmly bound together [2] and stored in the long-term memory system [10], when setting the recognition of an object representation in one sensory system at a lower threshold via selective attention to this sensory, the thresholds of representations of that object in other sensory systems is likewise decreased [34]. Thus, the task-irrelative auditory information that highly semantically matched with the visual target may be more preferentially and easily selected and processed, and it will enhance the identification of visual target under no load condition. Moreover, it has been proposed that when the time period between prime-target pairs that share a semantic relationship is shorter than 200 ms, the semantic priming processing for prime-target pairs of the same object is relatively automatic [49]. And once the semantically matched auditory stimulus has been selected and processed, its meaning may then unavoidably interact with any relevant visual information [64-65]. These findings would seem to suggest that the cross-modal attentional capture of semantically congruent AV stimuli may require minimal attentional resources, such that the presentation of a semantically congruent sound will enhance the recognition of the picture representation even when attentional resources is exhausted.

For semantically incongruent AV trials, it has been proposed that when attention is focused on visual modality, the concurrent and conflicting task-irrelevant auditory stream may capture bottom-up attention in a detrimental manner, and this cross-modal attentional capture mainly depend on the temporal correspondence[59]. Moreover, the conflicting nature of the task-irrelevant semantically incongruent sound was useless to the identification of the target picture [51], thus, in order to optimize the response, the semantically conflicting content between an ignored auditory and an attended visual stimulus should be not necessary to be processed, so that the temporal correspondence will make incongruent AV stimuli have better performance under no load condition. Moreover, semantically incongruent stimuli reflect the semantic violations of the multisensory semantic representation in long-term system [66-67], thus, it seems that these incongruent AV stimuli cannot be constructed as a coherent and stable whole and become very vulnerable to some top-down cognitive factors (such as attention and memory), for example, it has been found that incongruent AV stimuli should be overridden and quickly be forgotten [68-69]. We suspect that the instability of incongruent AV stimuli may also make it susceptible to the manipulations of attentional resources. Therefore, high load in which few attentional resources are left to process the ignored auditory stimulus, may greatly weaken the spatiotemporal audiovisual binding of incongruent AV stimuli, such that the auditory enhancement effect caused by semantically incongruent AV stimuli is eliminated by high load.

For semantically unrelated AV trials (animal images and white noise), we found AV integration effects for animal images and white noise only under no load condition, which were semantically unrelated and therefore no learned associations exist, hence, the auditory enhancement effect of semantically unrelated AV stimuli should be entirely due to temporal correspondence. Moreover, it is more difficult for the semantically unrelated AV trials, which were unlikely to cause a semantic violation, to be bound as a coherent and stable whole in comparison to incongruent AV trials, therefore, the auditory enhancement effect caused by semantically unrelated AV stimuli maybe also susceptible to the manipulations of attentional resources.

Our findings may also be explained based on the mechanism of temporal window of integration (TWI), i.e., the maximum temporal asynchrony between two different sensory events that allows them to be bound into a single, coherent multisensory event [70]. Even though it has been proposed that the boundaries of the temporal window of integration has strong plasticity, it can adaptively recalibrate after exposure to asynchronous stimuli in order to optimize performance according to specific task demands [71]. However, the multisensory integration can only occur over a given temporal interval range, if the temporal asynchrony between stimuli exceeds a certain temporal interval, it will be difficult to occur [72,73]. In the current research, as the degree of attentional load increases, the difficulty of RSVP visual search processing also increases, it will consumes more time to switch from the RSVP visual search task to the visual processing of AV integration task, therefore, increased load may result in the temporal interval between ignored auditory and visual target become larger, thereby increasing the difficulty of audiovisual binding. Furthermore, previous study has shown that the temporal window of integration is wider for semantically congruent stimuli than for incongruent, so that the AV integration of semantically congruent stimuli can be also occur even when asynchronous AV stimuli presented in relatively larger temporal interval [74,75]. This result is due to the fact that the temporal window of integration for semantically congruent stimuli can be widened to 300 ms with the help of conceptual short-term memory [51], given that this memory system may enables a person to retain these coherent multisensory congruent representations for an additional further 300 ms in order to achieve the participants' current task goal, while any incongruent and unrelated properties should be overridden and quickly be forgotten because they were useless to the identification of the target pictures [68-69]. Therefore, the AV integration of semantically congruent AV stimuli can effectively resist the impact of the increased temporal interval between stimuli due to the increased perceptual load, but the AV integration of semantically incongruent and semantically unrelated stimuli will be greatly affected by the increased load. Moreover, we found that increased load widened the temporal window of congruent AV integration, but not widen the temporal window of incongruent AV integration (Table 2 and 3). This furtherly suggests that increasing the attentional demands in our experiment might have resulted in an adaptive widening of the individual the temporal window of congruent AV integration that in turn led to the successful binding of the AV inputs, however, but this effect is not suitable for incongruent and unrelated inputs.

Furthermore, the enhancement effects in this study are presumably produced by the combination of spatiotemporal concordance and semantic congruency, and we cannot separate them in the present study, so it is not clear weather semantic congruency plays a decisive role in resisting the influence of interference task. Future research will focus on further separating the role of spatiotemporal concordance and semantic congruency in semantic AV integration processing to explore what factor makes the cross-modal integration of semantically congruent AV stimuli of common objects unaffected by limited attentional resource.

3.5. Conclusions

Our results demonstrated that attentional loads do not suppress the AV integration of semantically congruent AV stimuli but suppresses performance enhancements associated with semantically unrelated and conflicting AV stimuli under the condition of modality-specific selective attention. This study potentially indicates that, when attention is only focused on visual modality, the AV integration of semantically unrelated and incongruent AV stimuli depends on enough attentional resources, while the AV integration of semantically congruent AV stimuli occurs pre-attentively, requires little attentional resources. Thus, we further proposed that the strong semantic associations among AV stimuli plays an important role in resisting the effect of attentional load on the cross-modal integration processing.

Chapter 4 The Effect of Attentional Load on the Audiovisual Integration: When Spatial Congruence Matters

Summary

Even though prior studies have proposed that the cross-modal interaction of simple and arbitrarily paired audiovisual stimuli was also observed even when a particular modality was instructed to ignore, which may indicate that this cross-modal integration takes place automatically, regardless of attentional control. Nevertheless, it is not clear whether attentional resource is actually required for a cross-modal interaction of simple and meaningless audiovisual stimuli to obtain when auditory modality was instructed to ignore and whether attentional load would have different influence on the cross-modal interaction of simple AV stimuli presented at the same or different spatial positions. In the present study, we adopt an RSVP stream as the distractor task to manipulate different attentional loads: no attentional load, low attentional load, and high attentional load. Additionally, spatial congruency was controlled by presenting visual and auditory stimuli in the same or different locations. The current results revealed that spatially congruent auditory stimuli can provide a particularly effective means of improving the identification of visual targets even when they are engaged in high load condition, while spatially incongruent auditory stimuli cannot enhance the identification of visual targets under high load condition.

4.1 Background

When hiking in the forest, in order to locate a potential animal threat more quickly and accurately, using the combination of auditory and visual input information at the same time, as compared to using only auditory or visual information, will probably improves our ability to interact with the environment, allowing for faster detection and more accurate localization. Audio-visual (AV) integration is the phenomenon by which stimuli from visual and auditory sensory modalities can be integrated into a coherent representation to better perceive information [1]. Many studies have shown that attentional processes play a complex and multifaceted role in integrating input from different sensory modalities [5,18,20].

Attentional selectivity, the ability to remain focused on goal-relevant stimuli while ignoring others that are irrelevant to the current behavioral goal, is essential for any coherent cognitive function [9]. And it has been proposed that modality-specific selective attention plays a role in minimizing the cross-modal distraction from goal-irrelevant stimuli. Specially, modality-specific selective attention is one potential mechanism by which information from multiple sensory modalities could be filtered in order to amplifies the sensory neural responses for the selected goal-relevant signal and suppresses distraction responses for the goal-irrelevant information [50,76].

Some studies in which simple and arbitrarily paired bimodal stimuli were presented, have shown that attention can automatically spread from an attended visual stimulus to a task-irrelevant, simultaneously presented, unattended auditory stimulus, even when the two arise from different spatial locations [57-58]. Even though prior studies have proposed that even when a particular modality was instructed to ignore, the cross-modal interaction of simple and arbitrarily paired multisensory stimuli was also observed, which may indicate that this cross-modal integration takes place automatically, regardless of attentional control. Nevertheless, spare attentional resources may have contributed to this cross-modal interaction, due to attentional demands in this research failed to show any relatively inferior attentional modulation. Many studies have shown that participants can not completely ignore irrelevant visual inputs while attempting to respond selectively just to targets presented in relevant tactile modality [77]. Moreover, Load Theory posited that the level and type of load involved in the processing of goal-relevant information play a determined role in the processing of goal-irrelevant stimuli [22]; under high load condition, the relevant processing exhausts attentional resources such that irrelevant auditory stimuli can be successfully ignored [22,23]. Thus, if the ignored task-irrelevant distractors presumably have the 'special' quality that automatically influences the visual target detection, the integration of simple audiovisual stimuli under modality-specific selective attention should also occur even under high load. However, no study has explored the automatic quality of this integration under modality-specific selective attention by employing the dual task design in which different levels of endogenous attentional resources available is controlled. Therefore, it is not clear whether attentional resource is actually required for a cross-modal interaction of simple and meaningless audiovisual stimuli to obtain when auditory modality was instructed to ignore.

Spatial congruence was the locational correspondence of incoming signals from different sensory channels [10], and it can furtherly contribute multisensory information to produce facilitation effect [7,78]. Spence et al (2004) have found that spatial congruency has an impact on cross-modal visual-tactile integration processing (ignore visual) [11]. Furthermore, attention may have different effects on the processing of audiovisual stimuli presented at the same (an early influence of attention) or different (later attentional modulations) spatial positions [79-80]. It has been indicated that the spatial congruence of bimodal audio-tactile cue plays an important role in helping resist the interference of attentional load [13]. Specifically, when bimodal stimuli are presented in the same location, they effectively attract spatial attention even under high attentional load [14], if the location of tactile stimuli is neutral to auditory stimuli, the bimodal stimuli will not produce a spatial cuing effect or capture attention regardless of attentional load [13]. However, it is not at all clear whether attentional load would have different effects on the cross-modal interaction of audiovisual stimuli presented at the same or different spatial positions in the setting of modality-specific selective attention.

Thus, the aim of the present study was to explore whether the cross-modal interaction of AV stimuli can occur automatically, is not restricted by limited attentional resource in the setting of focused visual attention, and whether attentional load would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions. In the present study, we used a dual task paradigm to resolve these questions. And we will adopt an RSVP stream as the distractor task to manipulate different attentional loads: no attentional load (no load), low attentional load (low load), and high attentional load (high load). The participant was presented with an RSVP stream and either asked to ignore it (no load), detect infrequent yellow letters (low load), or detect infrequent white numbers (high load). Similar RSVP streams have been utilized to construct different level of perceptual loads in previously published dual task studies [81]. In the cross-modal interaction task, participants were instructed to respond to a specific image (black－white checkboard with two black dots) while ignoring all sounds (i.e., pure tone and white noise). And spatial congruency was controlled by presenting visual and auditory stimuli in the same or different locations.

4.2 Methods

4.2.1 Subjects

A total of 19 volunteers (five females, mean age of 25 years) participated in this study. The participants reported normal or corrected-to-normal hearing and vision. All participants provided written informed consent, and the study procedures were approved in advance by the ethics committee of Okayama University.

4.2.2 Stimuli

All study procedures were completed in a dimly lit, electrically shielded and sound-attenuated room, specifically, a laboratory room at Okayama University, Japan. Each participant positioned his or her head on a chin rest. All visual stimuli were presented on a 24-inch VG 248LCD monitor (made by ASUS, Taiwan) with a screen resolution of 1920×1080 and a refresh rate of 144 Hz set at a viewing distance of 57 cm from the participant. Auditory stimuli were presented through speakers located on the central monitor. Additionally, two speakers (Harman/Kardon HK206, frequency response: 90-20,000 Hz) were used to present the auditory stimuli. MATLAB software (R2014b, MathWorks, MA, Psychtoolbox-3) was used to present the experimental stimuli and record the participants' responses.

The visual target stimulus was a black and white checkerboard image with two black dots contained within the white checkerboard (52×52 mm, with a visual angle of 5 ˚), which was presented on a black background on a 24-inch computer monitor positioned 58 cm in front of the participant's eyes. The task-irrelevant visual stimulus was a black-white checkerboard image (Figure 1). All visual stimuli (V) were presented on the lower left or lower right quadrant of the screen for 150 ms (at a 12 ° visual angle to the left or right of the center and a 5-degree angle below the central fixation). The auditory target stimulus was a 1000-Hz white noise and the task-irrelevant auditory stimulus was a 1000-Hz sinusoidal tone. The auditory stimuli (A) were presented randomly to the left or right ear through earphones at approximately 75 dB SPL for 150 ms duration (10 ms of rise or fall cosine gate). The audiovisual stimuli (AV) were presented through a combination of the visual and auditory stimuli in the same or different spatial location (Figure 17).

Under each attentional load condition, AV integration task consisted of 160 V stimuli, 160 spatially congruent AV stimuli (AV_Con), and 160 spatially incongruent AV stimuli (AV_Incon). For each type of stimulus, 50% of trials presented the target stimulus. Thus, there were 6 different stimulus categories (trial types) consisting of the combination of stimulus modality (3 levels: V, AV_Con and AV_Incon) and stimulus type (2 levels: targets or standards). In addition, for the spatially incongruent AV stimuli (AV_Incon), half were presented the visual stimuli on the left and the auditory on the right, and the other half were presented the visual stimuli on the right and the auditory on the left. V and AV_Con stimuli were presented in either the left or the right hemispace with equal probability.

The stimuli in the RSVP task consisted of an S or a 5 which are coloured red, green, yellow, blue, purple or turquoise. In each trial, a stream of seven coloured characters (each subtending 2.3°×1.1°) were presented centrally one at a time.

4.2.3 Experimental design and procedure

The factorial design had two within subject factors: Stimuli type (V, AV_Con and AV_Incon), and Attentional load (no load, low load, high load).

In the present study, we employed a dual-task design to explore whether semantic congruency modulates the effects of attentional load on AV integration. First, we controlled for spatial congruency in the AV integration task; the spatially congruent/incongruent stimuli comprised visual targets presented along with either congruent or incongruent white noise. Second, we adopted the rapid serial visual presentation (RSVP) task used in Lunn et al (2019) as the distractor task to impose different levels of attentional load as follows: no load, low load, and high load [81]. Specifically, the participants simultaneously performed the AV integration task and a distractor task that required them to search a central RSVP stream for either a red letter (low load) or two different coloured letters (high load) in a series of different coloured characters. And participants were instructed to ignore the presented RSVP stream under the no-load condition [30].

Our study included three attentional load condition types by adopting an RSVP task, namely, no load, low load, and high load. In the low load condition, the AV integration task and RSVP task are displayed simultaneously [81] (see Figure 1). Each trial began with a central fixation cross presented for 400 ms, followed by a stream of seven coloured characters, presented centrally one at a time. And these coloured characters were continuously and sequentially displayed, being randomly replaced every 100 ms. Part of the task was to monitor a central stream of characters which were either an S or a 5, and could be coloured red, green, yellow, blue, purple or turquoise. Under the low load condition, the target of this RSVP task was either a red 5 or S (any red character). Specifically, the RSVP streams in each trial had a 50% probability of a target being present in each trial for all attentional-load conditions. The target of the RSVP task was presented with equal probability in the first through seventh positions in the stream. With respect to the targets of RSVP task, participants were asked to respond at the end of each trial, that is, after the interface of the red gaze point appeared, the subjects were to press the “J” button within 1000 milliseconds if they observed a target during the RSVP task.

In addition to the central RSVP task, participants were asked to monitor for peripheral visual targets which appeared on each trial, presented to the left or right of the central stream, and ignore of auditory stimuli. The peripheral stimuli were randomly presented during each trial for a period of 150 ms. For this AV integration task, Visual-alone, AV Congruent, AV Incongruent peripheral stimuli were presented equiprobably and in pseudorandom order to limit predictability. Participants were instructed to press the ‘3’ button if the visual target stimuli (black－white checkboard with two black dots) presented on the right hemispace and the ‘1’ button if presented on the left hemispace as rapidly and accurately as possible. Of note, the target stimuli were the pictures of black-white checkboard with two black dots or the AV stimulus including visual targets regardless of the accompanying sound (see Figure 1). A blank interface (1000 ms) was presented to ensure sufficient time to respond to peripheral visual targets.

In the high-load condition, the target was either a green 5 or yellow S, the other requirements were the same as those under the low-load condition, notably, because the task of searching for two different coloured characters in a series of letters (high load) requires more cognitive resources than the task of searching only for a specific colour (red) under the low-load condition. In this way, by increasing the difficulty of the distractor task, we can control the attentional resources that can be utilized by AV integration processing.

In the no-load condition, although the rapid serial visual presentation (RSVP) task was also displayed simultaneously with AV task, participants were asked to ignore of the RSVP task [30]. Participants did not need to judge the RSVP task, only to determine whether peripheral visual stimuli presented was a black-white checkboard with two black dots.

The experiment included 4 blocks of 144 trials each under each load condition, and each block lasted approximately 7 min. Thus, it takes about 28 min for each load condition. Participants were permitted to take breaks between blocks. In addition, each load condition was completed in a separate block, and the order in which participants completed the load condition blocks was randomized and counterbalanced across participants. Before the experiment was officially started, all participants engaged in a practice experiment with 30 trials to ensure that they correctly understood the experimental procedures and responded correctly to the different tasks.

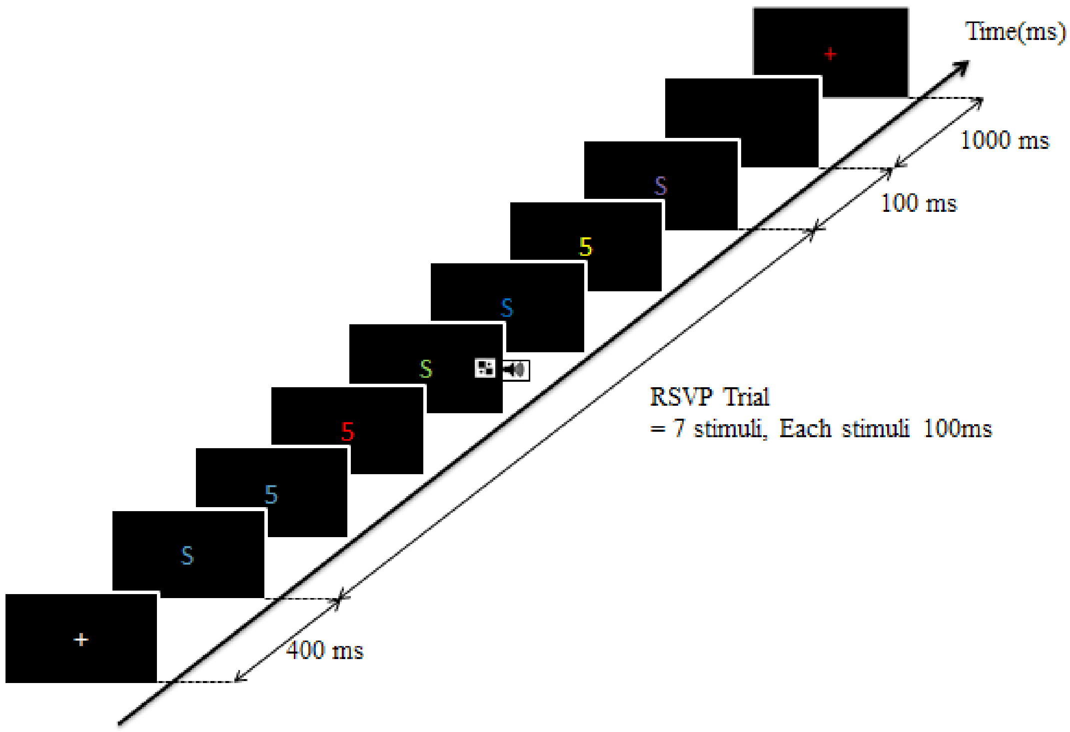

Figure 17. *A schematic representation trial in which both audiovisual integration task and the RSVP task were run simultaneously. The participants must judge whether there was a visual target (ignore white noise or pure tone) in the lower left or right corner of the screenwhile ignoring the RSVP streams (no load), reporting red S or 5 (low load), or reporting green 5 or yellow S (high load). Each trial began with a central fixation cross (400 ms), followed by a stream of seven characters (letters or numbers), which were sequentially presented with random replacement every 100 ms, while visual stimuli of audiovisual integration (150 ms) was randomly presented alongside the first to fifth letter of the RSVP streams. Participants should respond as soon as possible to the target picture by pressing the "1" (left) or "3" (right) key, and they were asked to press the "2" key for the target of the RSVP task when the red fixation point appeared (1000 ms).*

4.2.4 Data analysis

(1) Response time and accuracy analysis of the AV integration task

Participants' median accuracies and response times (RT) were calculated for the stimuli of different modalities under different attentional load condition after removing RT outliers (2.2%). Incorrect trials and trials with response times shorter than 200 ms or longer than 1000 ms were removed. To determine whether increased attentional loads have different effect on the cross-modal interaction of AV stimuli which have different spatial location, a repeated-measures analysis of variance with median response times as the dependent factor and Stimulus modality (Visual, Congruent AV, Incongruent AV) and Attentional load (No-load, Low-load, High-load) as independent factors was conducted. If the Shapiro-Wilk test was significant, percentage of accuracy was analyzed by nonparametric repeated-measures analyses of variance (ANOVA; the Friedman test). $P < 0.05$ was considered to be statistically significant. In addition, in each analysis, the degrees of freedom were corrected using the Greenhouse-Geisser correction when the Mauchly's test indicated that the assumption of sphericity had been violated.

(2) The analysis of auditory enhancement effect in the AV integration task

In addition, the auditory enhancement effects represent the percent change in performance for AV stimuli compared to visual stimuli [61], and the amount of improvement in the response times due to the addition of auditory stimuli were calculated by formula as follows:

$$\text{Auditory enhancement effect (\%)} = (RT_V - RT_{AV})/RT_V \times 100\% \quad (1)$$

RT_{AV} is the response times of the correctly detected AV stimulus and RT_V is the response times of the correctly detected visual stimulus. Furthermore, to test how attentional loads influenced the auditory enhancement effects produced by spatially congruent and incongruent AV stimuli, we conducted a repeated-measures analysis of variance on the auditory enhancement effects under different attentional load conditions.

(3) Cumulative Distribution Functions Calculation of AV integration task

To assess whether the magnitudes of the auditory enhancement effects were reduced by attentional loads and whether there was a difference under each load condition, we performed further analysis as follows. First, we subtracted the CDFs for responses to unisensory visual targets from the CDFs for responses to spatially congruent AV targets to obtain a measure of the auditory enhancement effects of spatially congruent AV stimuli [15,50]. The observed RTs of the congruent AV CDF was subtracted to the visual CDF at each time bin for each participant, to test for probability difference by performing planned pairwise comparisons under different attentional load conditions (No-load, Low-load, or High-load). Probability difference between congruent AV CDF and V CDF (i.e., $RT_{\text{Congruent AV}} < RT_V$) indicated the auditory enhancement that exceed statistical facilitation. Two-tailed P values were corrected for multiple comparisons using the Bonferroni correction, and differences are reported as significant at $p < 0.05$.

In addition, to test for differences in the positive area under the difference curve (i.e., the difference in probability of the semantically congruent AV CDF and the visual CDF for the RT range of 200 to 1100 ms) between attentional load, the positive area under the difference curve across the 10-ms time bins of the CDF of each load type of each participant was used in a repeated-measures ANOVA with the factor Attentional load (No-load, Low-load, High-load), followed by planned pairwise comparisons. In order to extract the positive area under the curve for each participant, the difference curve between the congruent AV CDF and the visual CDF was calculated for each participant. Next, all negative probabilities (no violation between congruent AV and V) were set to a value of zero and only the positive area under the curve was calculated for all participants. Furthermore, the greatest AV facilitation is defined as peak benefit, and the time spanning from the presentation of the target to the peak benefit is defined as the peak latency (see Fig. 17).

We subtracted the CDFs for responses to unisensory visual targets from the CDFs for responses to spatially incongruent AV targets to obtain a measure of the auditory enhancement effects produced by incongruent AV stimuli [15,50]. We adopted the same method as above to test for the positive area under the curve (i.e., the difference in probability of the spatially incongruent AV CDF and the visual CDF for the RT range of 200 to 1100 ms) between different attentional loads.

(4) Analysis of the influence of the distractor task

First, to check the RSVP performance to verify that participants accurately performed the distractor task (because they could have simply ignored it and only attended the primary task), the percentage of accuracy under different load conditions were analysed. A Shapiro-Wilk test was conducted to confirm the assumption of a normal distribution in low-load and high-load conditions. If the Shapiro-Wilk test was not significant, the repeated-measures ANOVA for comparisons between different load conditions were conducted. If the Shapiro-Wilk test was significant, we used the one-way nonparametric repeated-measures analyses of variance (ANOVA; the Friedman test) for comparisons. $P<0.05$ was considered to be statistically significant.

Second, we calculated the relative performance under the no load, low load and high-load conditions for all stimuli (visual, spatially congruent AV, spatially incongruent AV modalities) to explore whether attentional load significantly disrupted the RTs for the AV integration task.

4.3 Results

4.3.1 Response time and accuracy of AV integration

To determine how an increased attentional load interacted with spatial congruency to influence AV integration, we conducted repeated-measures ANOVA on median RTs using stimulus modality (visual, spatially congruent AV, spatially incongruent AV modalities) and attentional load (no-load, low-load, and high-load conditions) as factors. The results revealed a significant main effect of stimulus type, F (1.838, 33.075) = 54.529, $p < .001$, $\eta^2 = 0.752$. Participants responded significantly more slowly in visual target trials (*M* = 510.4 ms, *SD*= 12.1) than in spatially congruent AV (*M* = 484.8 ms, *SD*=11.4; t = 10.41, p = .001) or spatially incongruent AV (*M* = 492.3 ms, *SD*=11.7; t = 6.34, p = .001) target trials. Moreover, the main effect of the attentional load condition was also significant [F (1.877, 33.789) = 35.665, p < .001, η^2=0.665], with the participants responding more rapidly overall in the no-load condition (*M* = 399 ms, *SE*=7) than in the low-load (*M* = 463 ms, *SE*= 9; t = -5.35, p = .001) and high-load conditions (*M* = 496 ms, *SE*=11; t = -7.50, p = .001), as expected.

Interestingly, we also found a significant interaction between modality and attentional load [F (2.935, 52.833) = 3.981, p = .013, η^2= 0.181]. Post hoc subsidiary analyses with Bonferroni adjustment for multiple comparisons demonstrated that the median RTs for the spatially congruent AV target trials were significantly faster than were the visual RTs under each load condition [no-load condition: $t_{(18)} = 6.83$, p = .001; low-load condition: $t_{(18)} =5.43$, p = .001; high-load condition: $t_{(18)} =6.97$, p = .001] (see Figure 18A). Additionally, the median RTs for spatially incongruent AV trials were significantly faster than those for visual trials under the no-load, low-load conditions [no-load condition: $t_{(18)} = 6.47$, p = .001; low-load condition: $t_{(18)} = 3.24$, p = .014; high-load condition: $t_{(18)} = 2.73$, p = .041] (see Figure 18B).

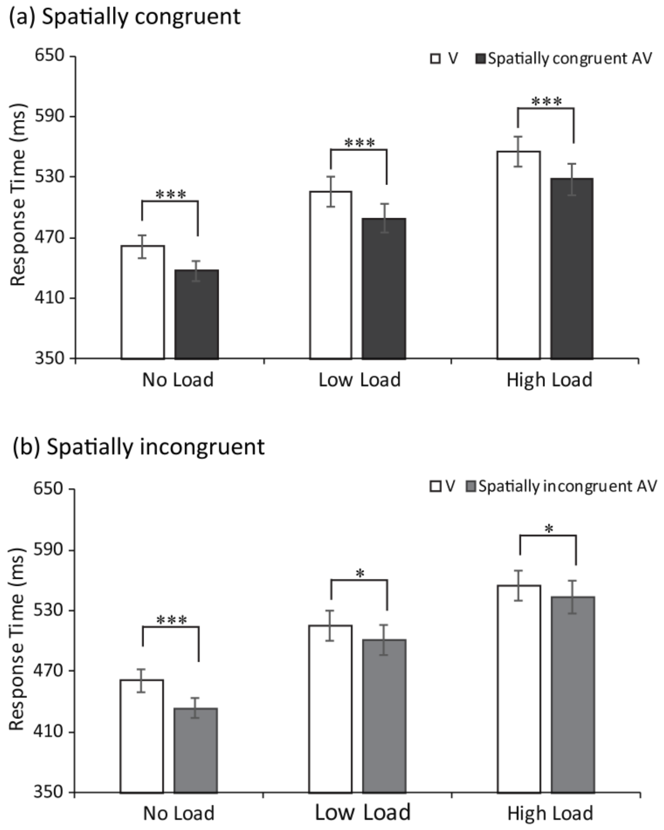

Fig. 18 *Median response times for the visual and spatially congruent AV trials (a), visual and spatially incongruent AV trials (b) in the animal identification task under different attentional load conditions. Error bars represent the standard errors of the means. ***p < 0.001, ** p < 0.01, * p < 0.05.*

4.3.2 Auditory enhancement effect in the AV integration task

Furthermore, to test whether the auditory enhancement effects produced by semantically congruent AV stimuli were influenced by attentional loads, a repeated-measures ANOVA was used. The repeated-measures ANOVA did not reveal a main effect of attentional loads [F (1.946, 35.032) = 0.009, p = .991, η^2=0.001]. Post hoc paired t-tests with Bonferroni correction showed that attentional loads did not significantly disrupt these auditory enhancement effects [no-load/low-load conditions: $t_{(19)}$ = 0.11, p = .99; no-load/high-load conditions: $t_{(19)} = 0.12$, $p = .99$; low-load/high-load conditions: $t_{(19)} = -0.01$, $p = .999$] (see Figure 19).

For the auditory enhancement effects of semantically incongruent AV stimuli, the repeated-measures ANOVA revealed a main effect of attentional loads [$F (1.897, 34.14) = 7.158$, $p = .003$, $\eta^2=0.285$]. Post hoc paired t-tests with Bonferroni correction showed that the auditory enhancement of incongruent AV trials was significantly larger in the high-load condition than in the no-load condition [$t_{(18)} = 3.11$, $p = .018$] and low-load condition [$t_{(18)} = 3.8$, $p = .004$], but there was no significant difference between the no-load and low-load conditions [$t_{(18)} = -0.5$, $p = .9$] (see Figure 19).

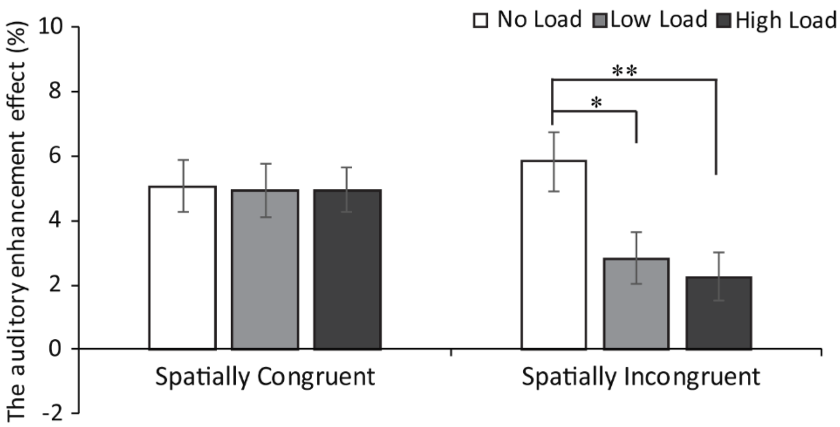

Fig. 19 *Comparison of the magnitudes of the median auditory enhancement effects (%) elicited by spatially congruent AV trials, by spatially incongruent AV trials under each load condition. Positive values indicate improvement, whereas negative values indicate impairment. ***p < 0.001, **p < 0.01, *p < 0.05.*

Table 5 *Median accuracy (%) and response times (RTs, ms) with standard deviations (SDs) for the target of each trial type under no -load, low-load, and high-load conditions.*

Stimuli	**No Load**		**Low Load**		**High Load**	
Type	**RT (SD)**	**Accuracy (SD)**	**RT (SD)**	**Accuracy (SD)**	**RT (SD)**	**Accuracy (SD)**
Visual	461.1 (49.4)	95.4 (3.1)	515.1(64.3)	89.5 (3.46)	555.0 (64.8)	77.5 (3.49)
Congruent AV	437.0 (42.3)	98.8 (1.84)	489.3 (61.2)	88.8 (2.56)	528.0 (68.0)	78.8 (4.16)
Incongruent AV	433.6 (43.9	95.6 (2.05)	500.4 (64.0)	86.8 (2.70)	543.0 (70.3)	75.8 (3.12)

4.3.3 Race modal of the AV integration task

To assess the effects of matching cross-modal promoters, we compared the unisensory visual CDF to the matching AV CDF under different attentional load conditions (Figure 20). A comparison between the unisensory visual target CDFs and semantically congruent AV target CDFs in each time bin revealed that auditory enhancement effects were observed under all load conditions ($p < .05$, paired t-tests, 2-tailed, Bonferroni corrected, Figure 20a). Specifically, the auditory enhancement effects occurred for 280 ms in the no-load condition (350-630 ms). In the low-load condition, this time window was 460 ms (340-800 ms), which was smaller than that in the high-load condition, had a window of 470 ms (330-800 ms).

Moreover, the positive area under the curve (the subtraction of the visual distribution from the congruent AV distribution) was compared between the different load conditions. The repeated-measure ANOVA revealed a main effect of attentional load condition [F (1.951, 35.12) = 1.098, p = .344, η^2 = 0.058]. Post hoc comparisons (Bonferroni corrected) showed that the positive area under the curve was not significantly larger in the no-load condition (*M* =13.4 ms, *SD* = 6.46) than in the low-load condition (*M*=17.0 ms, *SD* =9.7; $t_{(18)}$=-1.39, *p* = .544) and high-load condition (*M*=15.9 ms, *SD* =7.50, $t_{(18)}$ = -1.09, *p* = .869); also, there was no significant difference between the low-load and high-load conditions ($t_{(18)}$ = 0.43, p = .99) (see Fig. 20, right panel).

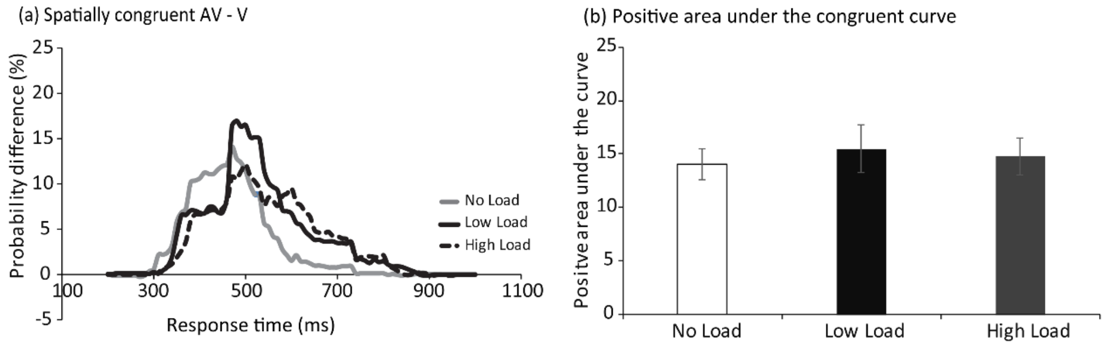

Fig. 20 *(a) The subtraction of the visual distribution from the spatially congruent AV target distribution across the full range of response times (RTs) for no-load (solid gray line), low-load (black dotted curve), and high-load (black dashed curve) conditions. Significant violations are indicated with a horizontal bar in the graph below the x-axis indicating the RT range in which the probabilities of congruent AV targets of consecutive time points are larger than the probabilities of visual targets. This result shows that significant auditory enhancement is present under all load conditions. (b) Median positive area under the curve (the subtraction of the visual distribution from the congruent AV distribution) in each load condition.*

To assess the effects of nonmatching cross-modal distractors, we compared the response distributions for unisensory visual trials to the response distributions for nonmatching AV trials under different perceptual conditions (Figure 21). The comparison between the unisensory visual target CDF and semantically incongruent AV CDF in each time bin showed an auditory interference effect; however, this auditory interference effect was only observed under the no-load condition and the low-load condition ($p < 0.05$, paired t-tests, 2-tailed, Bonferroni corrected, Figure 5a). Specifically, an auditory enhancement effect was observed at 300-640 ms in the no-load condition ($p < 0.05$), and a significant auditory enhancement effect was also found at 340-550 ms in the low-load condition ($p < 0.05$), but no auditory interference effect was found in the high-load condition. The negative area under the curve was also compared among the different load conditions (see Fig. 21, right panel).

The repeated-measures ANOVA revealed a main effect of load condition [$F_{(1.847, 32.254)} = 8.547$, $p = .001$, $\eta^2 = 0.322$]. Post hoc comparisons (Bonferroni corrected) showed that the positive area under the curve was not significantly larger in the no-load condition (*M* =15.5 ms, *SD* =8.0) compared to that in the low-load condition (*M*=8.7 ms, *SD* =6.1, $t_{(18)} = 3.35$, $p = .011$), but that in the no-load condition was significantly larger than that in the high-load condition (*M*=9.6 ms, *SD*=6.2, $t_{(18)} = 3.60$, $p = .006$). Moreover, there is no significant difference between the positive area under the curve in the low-load condition and in the high-load condition ($t_{(18)} = -0.51$, $p = .99$).

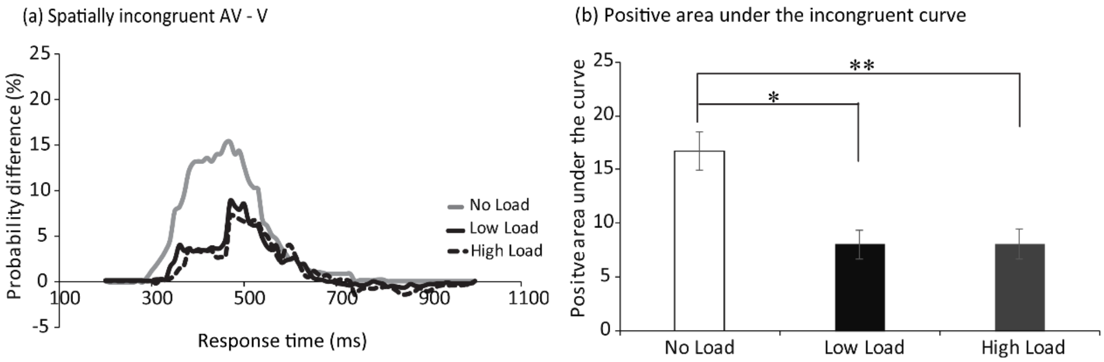

Fig. 21 *(a) The subtraction of the visual distribution from the spatially incongruent AV target distribution across the full range of response times (RTs) for no-load (solid gray line), low-load (black dotted curve), and high-load (black dashed curve) conditions. Significant violations are indicated with a horizontal bar in the graph below the x-axis indicating the RT range in which the probabilities of incongruent AV targets of consecutive time points are larger than the probabilities of visual targets. This result shows that significant auditory enhancement is only present under no-load and low-load conditions. (b) Median positive area under the curve (the subtraction of the visual distribution from the spatially incongruent AV distribution) in each load condition. Significant differences are indicated with an asterisk ($p < 0.05$).*

4.4 Discussion

The aim of the present study was to explore whether attentional load would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions. Our results support the claim that spatially congruent auditory stimuli can provide a particularly effective means of improving the identification of visual targets even when they are engaged in increased attentional load conditions (i.e., under conditions in which they had to perform two tasks at the same time; see [82]) (see Figures 18 and 20), while spatially incongruent auditory stimuli cannot enhance the identification of visual targets under high load condition(see Figures 18 and 21).

Consistent with the results of previous studies [58,83], we also found the improved detection of visual target stimuli accompanied by a task-irrelevant and temporally coincident auditory stimulus (whether or not it is spatially coincident with the visual target) under no load condition (see Figures 18 and 20). Buss et al., (2005) interpreted the attentional enhancement of the auditory processing as resulting from the grouping of the auditory and visual events, due to their temporal cooccurrence, into an audiovisual multisensory object [58]. Regarding the reason for the more improved detection of spatially misaligned audiovisual stimuli compared to unisensory visual targets, one possible might be that the enhancement effect produced by temporal alignment predominates the suppression effect of spatial misalignment in the specific case of audiovisual interactions.

Another possible might be that multisensory enhancement that is seen when both stimuli fall within the respective excitatory zones of the given neuron's receptive fields to each of the different sensory inputs, and the enhancement will be seen even with stimuli presented from different external locations if this excitatory zone is large [84-85]. Moreover, according to the spatial rule, only spatially coincident stimuli from different modalities are integrated, producing response enhancement, whereas spatially disparate stimuli produce response depression or else are not integrated, producing no interaction [86-87]. Many available evidences suggest that observation of the spatial rule in human multisensory perception/performance is very much task dependent [86,88,89,90]. Specifically, overt and covert spatial attentional orienting tasks are far more likely to give rise to evidence that is consistent with the spatial rule [88,91], whereas the data from tasks involving stimulus identification or temporal judgments rarely do. In addition, Girard et al. have proposed that when spatial information is task-irrelevant, multisensory integration of spatially aligned and misaligned stimuli is equivalent [91]. Furthermore, because we adopted stimulus identification task in our experiment (the spatial information is task-irrelevant), hence, simultaneous auditory stimuli arising from a different location also enhances the processing of visual targets.

Furthermore, we found attentional loads have different effect on the cross-modal integration of simple and arbitrarily AV stimuli presented at the same and different spatial positions (see Figures 18, 20 and 21). Our findings may be explained based on the "attentional load theory", which postulates that engaging attention in processing task-relevant stimuli with increased attentional loads substantially reduces and can even eliminate any neural signal related to potent task-irrelevant stimuli [22]. In other words, the processing of task-irrelevant information is reduced or eliminated when the perception of task-relevant stimuli under high load consumes all or most of the available capacity. Because participants are instructed to response to the visual target and ignore all the auditory stimuli in our experimental design, thus, ignored auditory stimuli are task-irrelevant, but visual stimuli are task relevant stimuli. Therefore, it's possible that under limited attentional resources the auditory input gained less attention, which resulted in the reduced processing of ignored auditory stimuli.

However, it has been proposed that the effects of perceptual load require clear spatial separation between the target and task-irrelevant distractor [22-23]. On the one hand, when both target and task-irrelevant distractor are parts of the same stimulus (e.g. a coloured word in the Stroop task), paying more attention to the target results in more attention to the task-irrelevant stimuli as well, high attentional load can increase Stroop interference [92]. Therefore, when the task-irrelevant auditory stimulus and visual target are presented at the same location (no spatial separation between the visual target and task-irrelevant auditory information), more attentional resources are also allocated to the processing of task-irrelevant auditory stimuli when participants are paying attention to the visual target. Hence, the task-irrelevant auditory stimulus which is spatially matched with visual stimulus can also enhance the identification of visual targets even when few attentional resources are left to process the task-irrelevant auditory stimuli.

On the other hand, it has been proposed that when there is a spatial separation of target and task-irrelevant distractor, high attentional load might further narrow the spatial attention window around the target space, resulting in the task-irrelevant stimuli to be effectively excluded from the processing range [22,92,93]. Therefore, when the task-irrelevant auditory stimulus and visual target are presented at different location (clear spatial separation between the visual target and task-irrelevant auditory information), limited attentional capacity under high load condition and visual search instruction might make participants' attention narrowing its focus to encompass just the visual target region, exclude task-irrelevant auditory stimuli outside it. Hence, when the task-irrelevant auditory stimulus is spatially mismatched with visual stimulus, it cannot influence or accelerate the identification of visual targets under high load condition.

4.5 Conclusions

The present study has explored whether attentional load would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions. Our results support the claim that spatially congruent auditory stimuli can provide a particularly effective means of improving the identification of visual targets even when they are engaged in increased attentional load conditions (i.e., under conditions in which they had to perform two tasks at the same time), while spatially incongruent auditory stimuli cannot enhance the identification of visual targets under high load condition. This study potentially indicates that, when attention is only focused on visual modality, the integration of spatially incongruent AV stimuli depends on enough attentional resources, while the AV integration of spatially congruent AV stimuli requires fewer attentional resources.

Chapter 5 General Conclusion and Future Projections

Summary

This thesis has investigated how semantic congruency interacts with attentional loads to influence the AV integration of common objects. Additionally, whether attentional loads would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions has also been evaluated. In this chapter, our findings are summarized below. Further, some future projections are included.

5.1 General Conclusions

The current thesis includes three experiment studies. The first experiment is the basis of the thesis, investigating how semantic congruency interacts with attentional loads to influence the AV integration of common objects behaviorally. The second experiment examined whether semantic association between AV stimuli modulates the effect of increased attentional loads on the AV integration of common objects in the setting of attended selectively to visual modality. In the third experiment, we explored whether attentional loads would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions.

Chapter 2 Describes how semantic congruency interacts with attentional loads to influence the AV integration of common objects by applying a dual-task paradigm. The dual-task paradigm reduces the attentional capacity dedicated to the main task (AV integration task in present study) because dividing attention between two concurrent tasks results in a decrease in behavioural performance relative to when only the main task is performed. We investigated these questions by examining AV integration under various attentional load conditions. Specifically, the participants simultaneously performed the AV integration task and a distractor task that required them to search a central RSVP stream for either a yellow letter (low load) or a white digit (high load); and the participants were instructed to ignore the presented RSVP stream under no load condition. The AV integration was assessed by adopting an animal identification task using unisensory (animal images and sounds) and AV stimuli (semantically congruent AV objects and semantically incongruent AV objects). The results confirmed that attentional loads did not attenuate the integration of semantically congruent AV objects. However, semantically incongruent animal sounds and images were not integrated (as there was no multisensory facilitation), and the interference effect produced by the semantically incongruent AV objects was reduced by increased attentional load manipulations. We further observed an asymmetric cross-modal interference effect supporting the visual dominance hypothesis; specifically, the auditory distractor effect was stronger than the visual distractor effect under all attentional-load conditions. These findings highlight the critical role of semantic congruency in modulating the effect of attentional load on the AV integration of common objects.

Chapter 3 Describes whether the cross-modal interaction of AV stimuli can occur automatically, is not restricted by increased attentional loads when attended selectively to visual modality, and whether semantic association between AV stimuli modulates the effect of increased attentional loads on the AV integration of common objects in the setting of attended selectively to visual modality. We manipulated the amount of available attentional resources by applying a dual-task paradigm and constructed three attentional load levels (no load, low load, and high load) by using a rapid serial visual presentation (RSVP) task. Semantic associations between AV stimuli were composed of animal pictures presented concurrently with either semantically congruent, incongruent or unrelated auditory stimuli. The results showed that attentional loads did not reliably alter the amount of the auditory enhancement effects caused by semantically congruent AV stimuli on this task. However, attentional loads disrupt the auditory enhancement effects of the semantically unrelated and incongruent AV stimuli. These findings suggested that the strong semantic associations between AV stimuli played an important role in withstanding the effect of attentional loads on cross-modal integration processing of modality-specific selective attention.

Chapter 4 Describes whether attentional loads would have different influence on the cross-modal interaction of simple and arbitrarily AV stimuli presented at the same or different spatial positions. We will adopt an RSVP stream as the distractor task to manipulate different attentional loads: no attentional load (no load), low attentional load (low load), and high attentional load (high load). Specifically, the participants simultaneously performed the AV integration task and a distractor task that required them to search a central RSVP stream for either a red letter (low load) or two different coloured letters (high load) in a series of different coloured characters. In the AV integration task, participants were instructed to respond to a specific image (black－white checkboard with two black dots) while ignoring all sounds (i.e., pure tone and white noise). And spatial congruency was controlled by presenting visual and auditory stimuli in the same or different locations. The results showed that significant audiovisual integration of spatial congruent AV stimuli occurred regardless of attentional load; however, increased attentional loads reduced the audiovisual integration of spatial incongruent AV stimuli. These findings highlight the critical role of spatial congruency in modulating the effect of attentional loads on the integration of simple and arbitrarily AV stimuli.

5.2 Future Projections

First, we have recently found that, in healthy young adults, the integration of semantically congruent audiovisual stimuli is relatively automatic, not attenuated by increased attentional loads [60], however, it remains unclear whether this can generalize to older adults and other special subjects. Specifically, although some studies found that a significant semantic audiovisual integration effect can be found in older adults [97], intersensory attention studies reported that attention plays an important role in semantic audiovisual integration [20], and older adults have some attentional deficits [94], and they are much more susceptible to irrelevant distractors [98]. Therefore, attentional deficits might be an important factor leading to the weaker audiovisual integration effect in older adults when irrelevant distractors were added (such as RSVP research task under increased attentional load conditions). Thus, although we have found that increased attentional loads do not reduce the integration of semantically congruent audiovisual stimuli in younger adults [60], it remains unclear how increased attentional loads influence the semantic audiovisual integration ability of older adults and other special subjects with attentional deficits, such as Parkinson's disease and Autistic disorder. Therefore, one future aim is to investigate the effect of increased attentional loads on the semantic audiovisual integration for special subjects.

It is important to address because aging and diseases can strongly affect the function of attention [96]. Understanding the relation between audiovisual integration and attention in these populations will help us elucidate the potential functional deficits in aging and disease on a fundamental process in perception (namely, audiovisual integration), which may help us devise potential interventions or identify therapeutic targets. Furthermore, potential alteration of semantic audiovisual integration under increased attentional loads may provide important basis for early clinical detection and potential rehabilitation of early aging and brain diseases. Therefore, it would of both scientific and clinical significance to investigate the effect of attentional loads on semantic audiovisual integration for special subjects.

In addition, although it has been found that attentional loads influence early neural processing of speech audiovisual integration [95], however, it remains unknown how increased attentional loads modulate the neural mechanisms underlying semantic audiovisual integration. Thus, we want to explore the effect of attentional loads on the neural mechanisms underlying semantic audiovisual integration for younger adults in the future research. According to a study conducted by Giard and Peronnet on the 'additive model' for multisensory integration, we subtracted the summation of ERPs evoked by unisensory-auditory and unisensory-visual stimuli from the ERPs evoked by audiovisual stimuli [100]. By comparing such topographical differences, another future aim was to clarify the mechanism of the effect of attentional load on semantic audiovisual integration by recording EEG signals from both unisensory-stimuli and semantically congruent audiovisual stimuli. Moreover, because older adults have attentional deficits [96] and distractor-suppression deficits [99], the effect of increased attentional loads on the semantic audiovisual integration for older adults may be different from younger adults. Therefore, another goal in future is to further clarify on the neural mechanism of the effect of attentional loads on the audiovisual integration for older adults and other special subjects.

Furthermore, based on the fact that different experimental findings may be obtained if we choose other alternative tasks as distractor tasks in future research, for example, if we choose other alternative tasks as distractor tasks in future research, we may obtain different experimental findings. For example, it has been proposed that when an object-based attention task is performed along with a spatial attention task, distinct attentional resources are required for the auditory and visual sensory modalities if a visual attentional load is induced [54]. Therefore, if a visuospatial task (i.e., a multiple object tracking task) was adopted as the visual distractor task, it selectively interfered with the visual discrimination task while the auditory discrimination performance was not affected. Furthermore, a question worthy of further investigation is whether multisensory integration can still occur even if the load task is multisensory.

Appendix

実験参加報告書

(実験者(記入)→実験参加者(記入:実験日)→実験者(受取)→GL→高橋)

【実験者 記入欄】

実験タイトル：

実験者：

実施年月日：　　年　　月　　日

実験時間：　　：　　～　　：　　(実験説明，休憩含む)

実験場所

実験項目(内容は簡単に，実験数に応じて記入)

実験１：

実験２：

実験３：

実験４：

実験５：

【実験参加者 記入欄】(表面，裏面とも記入し，署名する)

○現在の状態について，次の７段階で最も当てはまるスケールの数字に○印をしてください.

↓項目	当てはまらない		少し当てはまる		かなりあてはまる		非常に当てはまる
まぶたが重いと感じる	1	2	3	4	5	6	7
眠い	1	2	3	4	5	6	7
緊張している	1	2	3	4	5	6	7
どきどきしている	1	2	3	4	5	6	7
くつろいだ気分だ	1	2	3	4	5	6	7
ゆったりした気分だ	1	2	3	4	5	6	7
思考が鈍っている	1	2	3	4	5	6	7
認知の集中ができにくい	1	2	3	4	5	6	7
活力がみなぎっている	1	2	3	4	5	6	7
積極的な気分だ	1	2	3	4	5	6	7

やる気が出ない	1	2	3	4	5	6	7
何かすることに気乗りがしない	1	2	3	4	5	6	7
昨日は十分に睡眠をとった	1	2	3	4	5	6	7
昨日今日は暴飲暴食，深酒はしていない	1	2	3	4	5	6	7
健康状態は良好だ	1	2	3	4	5	6	7
実験中に不安に感じることがあった	1	2	3	4	5	6	7
次回も実験に参加したい	1	2	3	4	5	6	7
実験１の方法は理解できた	1	2	3	4	5	6	7
実験２の方法は理解できた	1	2	3	4	5	6	7
実験３の方法は理解できた	1	2	3	4	5	6	7
実験４の方法は理解できた	1	2	3	4	5	6	7
実験５の方法は理解できた	1	2	3	4	5	6	7
(以下必要に応じて実験者が追加)							
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
↑項目	当てはまらない		少し当てはまる		かなりあてはまる		非常に当てはまる

実験について，感想・コメントを１言(以上)書いてください．

【実験参加者 署名】

実験年月日：　＿＿＿＿年＿＿＿月＿＿＿日

署名(自署)：　＿＿＿＿＿＿＿＿＿＿＿＿＿＿

同意書への記入：　有　　　無

Publications

Journal Papers

[1] ***Qingqing Li,*** Qiong Wu, Yiyang Yu, et al. Semantic congruency modulates the effect of attentional load on the audiovisual integration of animate images and sounds. *I-perception*, in press, 2020. https://doi.org/10.1177/2041669520981096

[2] Lijuan Wang, ***Qingqing Li,*** Qiong Wu, Satoshi Takahashi and Jinglong Wu. The Categorical Relational Process Mechanism in Enactment Learning: Effects of Divided Attention and Categorical Cues. *Journal of Cognitive Psychology*, accepted, 2020. doi:10.1080/20445911.2021.1883032

[3] ***Qingqing Li***, Qiong Wu, et al. The Effect of Attentional Load on the Spread of Visual Attention across a Multisensory Object: When Semantic Associations matters. *Attention, Perception, psychophysics*, under revision, 2020.

International Conference Papers

[1] ***Qingqing Li***, Qiong Wu, et al. (2019, August). The Identification and Evaluation for Animal and Other Sounds: The Effect of Presentation Time. *2019 IEEE International Conference on Mechatronics and Automation* (*ICMA 2019*), pp. 874-879.

[2] Wu Wang, Han Ruiqi, R., Luo Yuluo, Wu Zehua, Jin Yin, Qingqing Li, & Li Bing. (2018, August). The Mediating Role of Self-Efficacy between Neuroticism and Procrastination among Undergraduates. *2018 IEEE International Conference on Mechatronics and Automation (ICMA 2018)*, pp. 67-71.

Conference Papers in Japan

[1]. **李 青青**，呉 瓊，呉 鳳侠，余 家斌，高橋 智，呉 景龍，江島義道: 視覚二重課題遂行時の聴覚刺激による促進効果の検討，日本生体医工学会中国四国支部大会．Ⅰ-3, p.20, (2018.10.27).

Acknowledgements

Firstly, I would like to express my sincerely gratitude to Prof. Jinglong Wu for the continuous support during my Ph.D studies. Prof. Jinglong Wu helped me in all my research design and writing of this thesis. I could not complete my study of doctor course and finish this thesis successfully without his enlightening instruction, impressive kindness and patience. I am very grateful to Prof. Jinglong Wu for his valuable comments during the submission of my paper. There is no doubt that Prof. Jinglong Wu is the leader of my scientific research and thesis writing. Thanks to Prof. Jinglong Wu for giving me guidance during the very confused and helpless submission process, my paper can be accepted smoothly. His diligence gives me power not only during my present PhD scours, but also in my future research working.

Secondly, I also want to express my sincerely gratitude to my supervisor Prof. Satoshi Takahashi. His rigorous scientific research attitude, serious and rigorous working style, as well as the concern and responsibility for the students' academic work will always be a role model for my future research work. I got a lot of comments for Prof. Satoshi Takahashi during I make my study plan, conduct my experiments, prepare for conference papers, write published papers, prepare for graduation announcement and this thesis. During the PhD scours, Prof. Satoshi Takahashi also greatly supports me when I applied the scholarships.

Thirdly, I also want to express my sincere gratitude to Dr. Qiong Wu, who provided very valuable advice for revising my study plan, conducting experiments, writing the dissertation, proposing kindly manuscript review and encouragement. In addition, Dr. Qiong Wu also helps me a lot for my daily life, let my life in Japan much easier.

Fourthly, I would like to thank the rest of my thesis committee: Prof. Gofuku and Prof. Yokohira for their insightful comments and encouragement, but also for the question which incented me to widen my research from various perspectives.

Last but not the least, I would also like to express my sincere thanks to Prof. Yoshimichi Ejima, who provided me a lot of comments during I write my papers. Without his precious support, it would not be possible to conduct this thesis successfully. I also thank the students and staff in the Wu lab. Without their cooperation, I cannot imagine how I could have finished my experiments. I sincerely thank all those who contributed to my experiment, my paper, and dissertation. And I thank my fellow lab mates in for the stimulating discussions and for all the fun we have had together during these years. I would like to thank my parents, my sisters and friends for supporting me spiritually throughout writing this thesis and my life in general.

References

1.Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, *11*(5), 473-490.

2.Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault Jr, T. J., ... & Schroeder, C. E. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience*, *31*(10), 1713-1720.

3.Beauchamp MS. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Current opinion in neurobiology 2005; 15 (2):145-153.

4.Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta psychologica*, *134*(3), 372-384.

5.Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain research*, *1242*, 136-150.

6.Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences*, *103*(21), 8239-8244.

7.Göschl, F., Engel, A. K., & Friese, U. (2014). Attention modulates visual-tactile interaction in spatial pattern matching. *PLoS One*, *9*(9), e106896.

8.Girard, S., Pelland, M., Lepore, F., & Collignon, O. (2013). Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Experimental brain research*, *224*(2), 275-285.

9.Zimmer, U., Roberts, K. C., Harshbarger, T. B., & Woldorff, M. G. (2010). Multisensory conflict modulates the spread of visual attention across a multisensory object. *Neuroimage*, *52*(2), 606-616.

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

11.Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial cuing studies of human crossmodal attention and multisensory integration. *Crossmodal space and crossmodal attention*, *277320*.

12.Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722-8726.

13.Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: when spatial correspondence matters. *Experimental Brain Research*, *195*(2), 261-272.

14.Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(6), 1311.

15.Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental brain research*, *158*(4), 405-414.

16.Murray, M. M., Michel, C. M., De Peralta, R. G., Ortigue, S., Brunet, D., Andino, S. G., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*, *21*(1), 125-135.

17.Xie, Y., Xu, Y., Bian, C., & Li, M. (2017). Semantic congruent audiovisual integration during the encoding stage of working memory: An ERP and sLORETA study. *Scientific reports*, *7*(1), 1-10.

18.Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400-410.

19.Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: bottom-up vs. top-down. *Multisensory Research*, *29*(6-7), 557-583.

20.Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*, 19.

21.Hartcher-O'Brien, J., Soto-Faraco, S., & Adam, R. (2017). A matter of bottom-up or top-down processes: The role of attention in multisensory integration. *Frontiers in integrative neuroscience*, *11*, 5.

22.Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75-82.

23.Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current directions in psychological science*, *19*(3), 143-148.

24.Bonato, M., Spironelli, C., Lisi, M., Priftis, K., & Zorzi, M. (2015). Effects of multimodal load on spatial monitoring as revealed by ERPs. *PloS one*, *10*(9), e0136719.

26.Plummer, P., & Eskes, G. (2015). Measuring treatment effects on dual-task performance: a framework for research and clinical practice. *Frontiers in human neuroscience*, *9*, 225.

27.Vroomen, J., Driver, J., & De Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources?. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(4), 382-387.

28.Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839-843.

29.Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*(3), 399-404.

30.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*, 1.

31.Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience*, *27*(30), 7881-7887.

32.Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cerebral cortex*, *18*(3), 598-609. doi:10.1093/cercor/bhm091

33.Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study. *Cerebral Cortex*, *14*(4), 452-465.

34.Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. J. (2007). Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*, *26*(2), 499-509.

35.Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research*, *194*(1), 91.

36.Mishra, J., & Gazzaley, A. (2012). Attention distributed across sensory modalities enhances perceptual performance. *Journal of Neuroscience*, *32*(35), 12294-12302.

37.Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*(2), 174.

38.Wagenmakers, E. J., Verhagen, J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for Bayesian hypothesis testing in psychological science. *Psychological science under scrutiny: Recent challenges and proposed solutions*, 123-138.

39.Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, *5*, 781. doi:10.3389/fpsyg.2014.00781

40.Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331-343.

41.Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive psychology*, *14*(2), 247-279.

42.Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics*, *78*(3), 723-735.

43.Van der Stoep, N., Van der Stigchel, S., Nijboer, T. C. W., & Spence, C. (2017). Visually induced inhibition of return affects the integration of auditory and visual information. *Perception*, *46*(1), 6-17.

44.Van der Stoep, N., Van der Stigchel, S., & Nijboer, T. C. W. (2015). Exogenous spatial attention decreases audiovisual integration. *Attention, perception, & psychophysics*, *77*(2), 464-482.

45.Hugenschmidt, C. E., Mozolic, J. L., & Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport*, *20*(4), 349.

46.Otto, T. U. (2019). RSE-box: an analysis and modelling package to study response times to multiple signals. *The Quantitative Methods for Psychology*.

47.Zimmer, U., & Macaluso, E. (2007). Processing of multisensory spatial congruency can be dissociated from working memory and visuo-spatial attention. *European Journal of Neuroscience*, *26*(6), 1681-1691.

48.Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing research*, *258*(1-2), 134-142.

49.Sachs, O., Weis, S., Zellagui, N., Huber, W., Zvyagintsev, M., Mathiak, K., & Kircher, T. (2008). Automatic processing of semantic relations in fMRI: neural activation during semantic priming of taxonomic and thematic categories. *Brain research*, *1218*, 194-205.

50.Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental brain research*, *184*(1), 39-52.

51.Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389-404.

52.Keitel, C., Maess, B., Schröger, E., & Müller, M. M. (2013). Early visual and auditory processing rely on modality-specific attentional resources. *Neuroimage*, *70*, 240-249.

53.Talsma, D., Doty, T. J., Strowd, R., & Woldorff, M. G. (2006). Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology*, *43*(6), 541-549.

54.Wahn, B., & König, P. (2017). Is attentional resource allocation across sensory modalities task-dependent?. *Advances in Cognitive Psychology*.

55.Chan, J. S., & Newell, F. N. (2008). Behavioral evidence for task-dependent “what” versus “where” processing within and across modalities. *Perception & Psychophysics*, *70*(1), 36-49.

56.Wahn, B., & König, P. (2016). Attentional resource allocation in visuotactile processing depends on the task, but optimal visuotactile integration does not depend on attentional resources. *Frontiers in integrative neuroscience*, *10*, 13.

56. Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, *61*, 208-224.

57.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*(3), 679-690.

58.Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences*, *102*(51), 18751-18756.

59.Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter?. *Cerebral Cortex*, *20*(1), 109-120.

60.Li, Q., Wu, Q., Yu, Y., Wu, F., Takahashi, S., Ejima, Y., Yang, J., & Wu, J. (2020). Semantic congruency modulates the effect of attentional load on the audiovisual integration of animate images and sounds. *i-Perception*, *11*(6), 1-24.

61.Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and hearing*, *26*(3), 263-275.

62.Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental brain research*, *184*(1), 39-52.

63.Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog’s meow: asymmetrical interaction in cross-modal object recognition. *Experimental Brain*

Research, *193*(4), 603-614.

65.Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of experimental psychology: General*, *133*(3), 339.

66.Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: bottom-up vs. top-down. *Multisensory Research*, *29*(6-7), 557-583.

67.Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience*, *27*(30), 7881-7887.

68.Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. *Fleeting memories: Cognition of brief visual stimuli*, 13-46.

69.Potter, M. C. (1993). Very short-term conceptual memory. *Memory & cognition*, *21*(2), 156-161.

70.Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598-607.

71.Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517.

72.Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*, *203*(2), 381-389.

73.Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience*, *29*(39), 12265-12274.

74.Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PloS one*, *4*(5), e5664.

75.Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & psychophysics*, *69*(5), 744-756.

76.Kastner, S., McMains, S. A., & Beck, D. M. (2009). Mechanisms of selective attention in the human visual system: Evidence from neuroimaging.

77.Pashler, H. E. (Ed.). (1998). *Attention*. Psychology Press.

78.Goldring, J. E., Dorris, M. C., Corneil, B. D., Ballantyne, P. A., & Munoz, D. R. (1996). Combined eye-head gaze shifts to visual and auditory targets in humans. *Experimental brain research*, *111*(1), 68-78.

79.Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177-180.

80.Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722-8726.

81.Lunn, J., Sjoblom, A., Ward, J., Soto-Faraco, S., & Forster, S. (2019). Multisensory enhancement of attention depends on whether you are already paying attention. *Cognition*, *187*, 38-49.

82.Santangelo, V., Van der Lubbe, R. H., Belardinelli, M. O., & Postma, A. (2008). Multisensory integration affects ERP components elicited by exogenous cues. *Experimental Brain Research*, *185*(2), 269-277.

83.McDonald, J. J., Teder-Sälejärvi, W. A., Russo, F. D., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*, *15*(1), 10-19.

84.Kadunce, D. C., Vaughan, J. W., Wallace, M. T., Benedek, G., & Stein, B. E. (1997). Mechanisms of within-and cross-modality suppression in the superior colliculus. *Journal of neurophysiology*, *78*(6), 2834-2847.

85. Wallace, M. T., & Stein, B. E. (2007). Early experience determines how the senses will interact. *Journal of neurophysiology*, *97*(1), 921-926.

86.Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, *1296*(1), 31-49.

87.Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental brain research*, *160*(3), 273-282.

88.Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial cuing studies of human crossmodal attention and multisensory integration. *Crossmodal space and crossmodal attention*, *277320*.

89.Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., & Molholm, S. (2011). Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *Journal of Neuroscience*, *31*(27), 9971-9981.

90.Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1063.

91.Girard, O., Mendez-Villanueva, A., & Bishop, D. (2011). Repeated-sprint ability—Part I. *Sports medicine*, *41*(8), 673-694.

92. Chen, Z. (2003). Attentional focus, processing load, and Stroop interference. *Perception & Psychophysics*, *65*(6), 888-900.

93. Fernandez-Duque, D., & Johnson, M. L. (2002). Cause and effect theories of attention: The role of conceptual metaphors. *Review of general psychology*, *6*(2), 153-165.

94. Handy, T. C., & Mangun, G. R. (2000). Attention and spatial selection: Electrophysiological evidence for modulation by perceptual load. *Perception & psychophysics*, *62*(1), 175-186.

95. Fraser, S., & Bherer, L. (2013). Age-related decline in divided-attention: From theoretical lab research to practical real-life situations. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(6), 623-640.

96. 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*, 727.

97. Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of aging*, *27*(8), 1155-1163.

98. Williams, D. R., & Collins, C. (2016). Racial residential segregation: a fundamental cause of racial disparities in health. *Public health reports*.

99. Quigley C, Andersen S. K, Schulze L, Grunwald M, Muller M. M. (2010). Feature-selective attention: Evidence for a decline in old age. *Neuroscience Letters*, 474 (1), 5–8.

100. Giard MH, Peronnet F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, 11 (5), 473-490.