Title of Thesis

Study on audiovisual integration and cross-modal working memory training in young and older adults

September 2023

Ao Guo

Graduate School of Interdisciplinary Science and Engineering

in Health Systems

(Doctor’s Course)

OKAYAMA UNIVERSITY

Abstract

Audiovisual integration refers to the phenomenon in which individuals effectively perceive visual and auditory information presented simultaneously as a unified and coherent perceptual experience. It can enhance individuals' information recognition, resulting in more accurate and faster judgments of the external world. However, the topic of audiovisual integration in older and young adults remains a subject of debate and requires further exploration. Early research primarily focused on the effects of audiovisual integration on detection tasks, while its impact on higher-level cognitive processes, such as working memory, requires further investigation. Due to differences in perceptual and cognitive abilities between older and young adults, it is unclear whether audiovisual integration has different effects on working memory in these two groups. In addition, there is emerging evidence that working memory can be improved by training. Most working memory training tasks tend to be restricted to a single modality. However, the combined information from visual and auditory modalities could lead to audiovisual facilitation and enhance our ability to perceive the world. Therefore, we tested whether audiovisual working memory training led to training effects on working memory and transfer effects on perceptual processing. If audiovisual integration has a positive impact on working memory, can we improve working memory capacity through audiovisual working memory training and transfer those benefits to other related abilities across different age groups? This remains an open question.

This study consisted of three experiments. Experiment 1 aimed to investigate the

neural and behavioral differences between older and young individuals in the visual modality, auditory modality, and audiovisual modality by manipulating a simple detection task. Additionally, it sought to elucidate the differences in audiovisual integration between older and young individuals. The results revealed that older adults exhibited significantly greater audiovisual integration than young adults in terms of behavioral performance. Neurologically, older adults demonstrated earlier audiovisual integration in the early stages (110-140 ms, 140-170 ms) compared to young adults (190-250 ms), but delayed integration in the later stages (350-380 ms) compared to young adults (300-360 ms).

In Experiment 2, we compared the audiovisual n-back task to the visual n-back and auditory n-back tasks to determine whether working memory in the audiovisual condition could lead to better performance. The results showed that reaction time in the audiovisual condition was faster than that in the visual and auditory conditions across both young and older adults, indicating that audiovisual integration has a positive impact on working memory. Considering the trade-off between accuracy and reaction time, the inverse score (IES) was conducted. The results showed that the IES scores of older individuals are lower under audiovisual conditions compared to their scores under visual-only and auditory-only conditions. The IES scores of young individuals under audiovisual conditions are lower than their scores under auditory-only conditions. Moreover, greater visual contribution was found in older adults, especially at 3-back condition.

In Experiment 3, we investigated whether audiovisual working memory training

could induce training effects and transfer effects. For young adults, we found that working memory performance improved, as reflected by accuracy and response time across the 1-back, 2-back, and 3-back levels after training. The ERP (event-related potential) results showed that the P300 component was enhanced in the frontal and central regions across all conditions. Additionally, the N200 component was also enhanced in the central region in the 3-back level. In terms of neural oscillations, alpha oscillations and theta oscillations were enhanced after training at 250-300 ms. Transfer effects were also observed in young adults, as reflected by greater audiovisual integration in the early stage (80-120 ms) in the training group. For older adults, the behavioral results showed that audiovisual n-back training led to a training effect on the 3-back level, which represents a relatively higher cognitive load. The ERP results showed that the N200 latency was earlier after training. In terms of neural oscillations, alpha oscillations were enhanced at 220-250 ms. For the transfer effect, greater audiovisual integration was found at 180-200 ms, indicating that audiovisual working memory training yielded transfer effects on audiovisual processing.

KEYWORDS

audiovisual n-back, training, audiovisual integration, ERPs, training effect, transfer effect

Table of Contents

Chapter 1 Introduction

1.1 Audiovisual integration

In the complex and ever-changing world, individuals usually receive different information from multiple modalities and process this information into a whole perception. For instance, when tasting food, visual, olfactory, and gustatory information are integrated into one perception, which may help us better identify the food. It seems that our brain has the capacity to integrate these cross-modal information into a unique and coherent perception [1], which is regarded as multisensory integration. Notably, previous researches have shown that 94% of information encoded in our brain were from visual and auditory modality, addressing their importance in higher-order cognitive processing [2,3]. Therefore, in the domain of multisensory integration, audiovisual integration has been recognized as a significant area of research. Numerous studies have demonstrated that audiovisual integration plays an important role in our daily life by facilitating and improving our perceptual processing. Specifically, audiovisual integration combines visual and auditory information, which may result in different responses at the behavioral and neural levels compared to single visual or auditory modalities [1,2]. This advantage is called redundant target effect (RTE). The RTE at the behavioral level is reflected by the fact that higher accuracy and lower reaction time in audiovisual condition is found compared to these indexes in visual and auditory condition. On the other hand, the RTE at the neural level indicates that the neural activity was reduced during the presentation of audiovisual stimuli when compared to the sum of visual and auditory neural activity, indicating that individuals can process the audiovisual stimuli with higher neural efficiency [3]. This leads to the question of whether audiovisual integration can occur when visual and auditory stimuli are combined. Essentially, audiovisual integration operates based on four principles. Firstly, to achieve optimal outcomes, the unimodal sensory stimuli should be presented with temporal and spatial congruency. The final factor is known as inverse effectiveness, suggesting that audiovisual integration is most effective when less intense or weak stimuli are presented.

1.1.1 Spatial-temporal factor

The spatial-temporal factor refers to the phenomenon that when visual and auditory stimuli are presented in the same temporal order and same spatial location, the effect of audiovisual integration is optimal [4]. This finding suggested that spatial and temporal factors are two keys for audiovisual integration. The ventriloquism effect, also known as the phonetic shift effect, is a visual-auditory illusion where, when we hear a sound while observing a visual stimulus in a different location, we tend to incorrectly perceive the sound source as coming from the same location as the visual stimulus [5]. Notably, by increasing the spatial and temporal separation between auditory and visual stimuli, the magnitude of the ventriloquism effect would be reduced [6]. For the spatial factor, some studies have shown that superior colliculus (SC) was associated with orientation. Multisensory neurons have multiple excitatory receptive fields for different sensory modalities, and these receptive fields are spatially aligned with each other. Therefore, when stimuli from two different sensory modalities fall within the space where their receptive fields overlap, the neuron is activated and defines these two stimuli as originating from the same source location [7]. The temporal factor demonstrates the temporal relationship between stimuli in different modalities which would be integrated. That is, stimuli from different sensory modalities are more likely to be integrated when they occur closer in time [7]. Some fMRI studies have different activation patterns in different temporal synchronization. Calvert et al. used fMRI technology in their research to investigate audiovisual integration in speech perception. They presented subjects with visual conditions (watching lip movements while reading a story), auditory conditions (listening to the story), and audiovisual conditions (both visual and auditory stimuli presented together). The audiovisual condition was further divided into two types, one with congruent visual and auditory stimuli and the other with incongruent stimuli that conveyed different information and were not synchronized in time[8]. Results showed greater brain activation in the left superior temporal sulcus was found in audiovisual condition compared to the visual and auditory conditions. However, such enhanced brain activation related to supra-additive response was observed only in the condition with audiovisual congruent stimuli [8].

Despite possible temporal discrepancies in the arrival of visual and auditory stimuli, the nervous system can process these stimuli within a relatively broad time window, allowing for multisensory integration to occur. This adaptability enables the brain to accommodate time differences between different sensory modalities and integrate this information throughout the entire time window, facilitating more efficient processing of information from various sensory modalities. Specifically, the temporal binding window for audiovisual integration refers to the specific time range during which visual and auditory stimuli need to occur simultaneously or in close temporal proximity to be perceived as occurring simultaneously by the perceptual system [9]. Notably, this window has plasticity across different age groups [9]. In the research of Bedard et.al., (2016), three tasks were used to measure the audiovisual processing including temporal order judgment (TOJ), simultaneity judgment (SJ) and stream illusion, across older and young adults. As expected, the study found that older adults faced challenges in temporal processing, as evidenced by a larger temporal binding window in tasks involving temporal order judgment (TOJ) and stream illusion. Conversely, the performance of simultaneity judgment (SJ) was similar between older and young adults. The researchers concluded that older adults exhibit an extended temporal binding window for TOJ and stream tasks, but the temporal binding window for SJ remains relatively unaffected [9].

1.1.2 Inverse effectiveness

Behavioral studies have demonstrated that the magnitude of audiovisual integration is inversely related to stimulus intensity, reflected by redundant target effect (RTE) [10]. The presentation of audiovisual stimuli with low-intensity condition leads mor robust audiovisual facilitation compared to that of high-intensity condition. The subsequent ERPs study have shown that audiovisual integration in early stage (40-60 ms) with a left posterior and right anterior regions were found particularly for stimuli with low-intensity condition [10]. Moreover, our recent studies even found that the inverse effectiveness varied in older and young adults. An audiovisual discrimination task was conducted in both aged groups. The results showed that greater audiovisual integration was found in older adults at 320-360 ms. Adversely, at 460-500 ms, attenuated audiovisual integration in the frontal, fronto-central central and centro-parietal regions was found in the older adults compared to the young adults. In addition, the authors found greater audiovisual integration under low-intensity condition than under high-intensity condition in older adults at 200-230 ms, showing that inverse effectiveness occurred. However, such an effect was not found in young adults. These results suggested that audiovisual integration and inverse effectiveness were dissociated with age, indicating that the neural mechanisms of audiovisual integration between older adults and young adults differed [11]. The furthered fMRI research revealed specific brain regions responsible for inverse effectiveness. In more details, two fMRI studies using audiovisual tool and speech stimuli presented at high-intensity and low-intensity conditions, the results demonstrated inverse effectiveness was found in superior temporal sulcus (STS). Although audiovisual tool and speech stimuli defined regions of interest were non-overlapping, the pattern of inverse effectiveness was the same for both types of stimuli across regions, indicating that the manner in which visual and auditory stimuli are integrated in STS is not specific to speech [12].

1.1.3 Neural mechanism of audiovisual integration

Early research on animals found that the region of superior colliculus was associated with audiovisual processing. It is acknowledged that the superior temporal sulcus (STS) located in the temporal lobe and the interparietal sulcus (IPS) have been involved in audiovisual processing [13]. Moreover, some primary cortex were also engaged in audiovisual processing including visual or auditory areas [14]. Audiovisual integration at the cellular level was also portrayed. In simple terms, audiovisual integration refers to the phenomenon where a neuron generates a significantly higher number of impulses when simultaneously presented with both visual and auditory stimuli compared to the number of impulses elicited by the most effective single stimulus (either visual or auditory). This increased neuronal response to audiovisual stimuli indicates that the brain's response is stronger when integrating information from different sensory systems compared to a response to a single stimulus. This phenomenon is considered the physiological basis of audiovisual integration in neuroscience research [7]. Notably, the study of our colleague also provide support for multiple stage of audiovisual integration in terms of temporal dynamic processing. That is, audiovisual integration occurs at both early processing stage (60-90 ms and 110-140 ms) and later processing stage (220-250 ms and 340-370 ms) [15].

Moreover, one of the central questions is that whether audiovisual processing leads to activation solely in primary cortices or also involves activation in higher-order cortices. This may involve exploring the stages of audiovisual integration processing, whether it occurs early, late, or at both early and late stages. Specifically, if audiovisual integration occurs in the early stages, the activated brain regions may primarily involve the primary cortices. If it occurs in the later stages, the activated brain regions may involve higher-order cortices. However, if audiovisual integration takes place across multiple stages, both primary and higher-order cortices may be involved in the activation process. Early-stage models suggests that audiovisual processing occurs simultaneously with early perceptual processing [16]. Some studies have supported this idea through ERPs research on audiovisual integration. Their results found that audiovisual integration could activate primary visual cortex as early as 40 ms after stimulus onset [14]. The functional significance of such early sensory integration remains to be determined [7]. In contrast, the late-stage model tends to view audiovisual integration as a higher-level cognitive process that typically occurs after the separate processing of visual and auditory inputs has been completed [16]. The concept has evolved from inflexible principles of early-stage or late-stage integration to a highly adaptable process encompassing multiple stages of integration [17]. By using sin-wave speech stimuli, Baart et al., (2014) identified two distinct audiovisual processing stages in the integration process. The first processing stage is dominated by the auditory N100 component and is not influenced by the speech mode. The second processing stage is dominated by the P200 component and is influenced by the presence of visual information, but it occurs only when participants are in the speech mode [18]. It is worth noting that, compared to simple stimuli, audiovisual integration in the speech mode has a later and longer latency of occurrence [10,14]. The early N100 component may be associated with automatic process and the later P200 is related to higher cognitive function called top-processing [17]. From the perspective of brain activation, research has also supported the idea that audiovisual integration involves multiple processing stages, with both top-down and bottom-up processes being implicated. Our recent study using ERPs have also supported the idea that audiovisual processing occurs at multiple stages. That is, audiovisual integration was found at 200-230 ms, 320-360 ms, 460-500 ms [11]. FMRI results have shown that simultaneous presentation of visual and auditory stimuli leads to enhanced activity in multiple brain areas involved in audiovisual processing [19]. In addition, a time window of integration model suggests that audiovisual processing includes two consecutive phases: an early stage of peripheral processing and a later stage of converging sub-processes. Particularly, in the first processing stage, visual and auditory information are processed separately. Then, within a specific integration time window, if both independent processing processes are completed, the second stage begins. In the second stage, the brain integrates the information from the visual and auditory modalities, leading to audiovisual integration [20]. The second hypothesis was associated with predictive coding theory, which demonstrated that our perception was determined by a trade-off between prediction based on prior experience accumulated over our life span and efficiency of sensory processing [21-23].

1.1.4 Audiovisual integration during aging

The degree of aging in the peripheral sensory organs and cortical structures responsible for higher-order cognitive processing significantly influences the processing of audiovisual stimuli. As individuals age, their perceptual systems and cognitive functions tend to decline to varying extents, including reduced visual and auditory sensitivity, decline in working memory capacity, and attentional abilities [24-27]. Additionally, with aging, changes occur in brain structures, such as reductions in gray and white matter volume and alterations in the activation levels of brain regions responsible for specific cognitive processes [13]. These various brain changes resulting from age-related factors may have an impact on audiovisual integration.

A considerable number of ageing studies have compared audiovisual integration between young adults and older adults. Some researchers employed a color-naming task with semantic stimuli under visual, auditory, and audiovisual conditions [28]. The results showed that the reaction time to audiovisual stimuli was accelerated in young and older adults, with a significantly greater audiovisual gain in older adults. Similarly, in a simple discrimination task with meaningless stimuli, enhanced audiovisual integration was also found in older adults [29]. In order to get a clearer and more complete understanding of the neural changes associated with audiovisual processing during aging, subsequent ERPs research revealed that the neural response to audiovisual stimuli in older adults was stronger than that in young adults [30]. The stronger neural response was further confirmed at medial prefrontal regions and inferior parietal regions 100 ms after stimulus onset in older adults [30]. These studies support the point that the enhanced audiovisual integration related to the activation of specific brain regions in older adults may offer a strategy to compensate for the deficiency of unimodal sensory processing [28]. The studies mentioned above regarding audiovisual integration in older adults mostly adopted a simple combination of auditory and visual stimuli and were based on non-speech paradigms. However, audiovisual integration in older adults was equivocal for speech perception, which was more complex and needed more cognitive processing [31]. Sommers et al. (2005) examined the age-related difference of audiovisual integration with speech stimuli, by testing the contribution of visual and auditory processing to audiovisual integration [32]. The results showed that the audiovisual gain in older adults was reduced compared to young adults. The authors further proposed that the capacity of encoding visual information in older adults was reduced, which may diminish the potential to comprehend auditory information [32]. However, some researchers reported contradictory findings that audiovisual speech processing did not deteriorate in older adults. That is, they can obtain equivalent audiovisual gain as young adults, with a larger reduction in the amplitude of N1 component in the audiovisual condition relative to the summed amplitude of auditory and visual conditions [3]. This neural alteration of older adults in audiovisual speech processing might be due to the accumulated experience, suggesting that older adults were more capable of extracting useful information from visual stimuli (lip movement) to predict the upcoming auditory stimuli (spoken utterance) [3].

In summary, older adults were inclined to obtain more audiovisual integration compared to young adults. Several hypotheses could be used to explain the increased audiovisual integration in older adults including top-down and bottom-up processing, compensation strategy and deficits in attentional control. Multisensory integration involves top-down and bottom-up processes. The former process can be affected by attention, memory, and expectation, while the latter is driven by stimulus salience. According to a review discussing the stages of multisensory integration, the authors suggest that multisensory integration is largely subserved by a top-down process [17]. Previous studies have shown that frontal and fronto-central regions are highly associated with cognitive function, such as top-down attention [33]. Nevertheless, during ageing, older adults exhibit a greater disadvantage over top-down attention due to the decreased inhibitory function [34]. This may offer an interpretation for older adults having enhanced audiovisual integration over these regions compared to young adults.

Secondly, the compensatory mechanism in the aging brain refers to an adaptive strategy where older adults, facing age-related cognitive decline and brain structural changes activate other brain regions or increase brain activity to compensate for functional deficits. This compensatory mechanism aims to help older adults maintain relatively higher levels of performance in cognitive tasks compared to young adults. During memory task, more brain activity in the prefrontal cortex were engaged in older adults, indicating that the greater activation in this brain region compared to young adults may compensate for the attenuated activity in visual processing regions [35]. Similarly, during audiovisual integration processes, older adults may show increased brain activity to compensate for deficiencies in sensory processing [11]. The effectiveness of the compensatory mechanism may vary across individuals and can be influenced by task demands and cognitive load. Under lower cognitive load levels, the compensatory mechanism may be effective. However, as tasks become more complex or demanding, cognitive resources may reach their limitations, leading to the fact that the compensatory effect becomes ineffective [36].

Thirdly, deficits in attentional control of older adults may result in a weak ability to suppress audiovisual information. Previous studies conducted a cued audiovisual discrimination task to explore the alteration of audiovisual integration under selective and divided conditions in different age groups [37]. The results not only showed that older adults had greater audiovisual integration under selective condition, but also found that audiovisual integration was still enhanced in older adults relative to young adults under divided conditions [37]. This might be because young adults could ignore modality-specific irrelevant information, further suppressing the audiovisual integration. As for older adults, they could not judge whether the information was irrelevant effectively, leading to enhanced audiovisual integration. Moreover, their subsequent study found neural evidence that older adults showed greater brain activity than young ones when they were involved in irrelevant information [38]. The greater audiovisual integration in older adults reflected that they were more distracted by the irrelevant information. This phenomenon was also found in other audiovisual perception tasks, such as the Mcgurk task or the sound-induced illusion task, in which the older group had a higher possibility of being influenced by the interaction between visual modality and auditory modality and was more vulnerable to yield the audiovisual effect [39,40].

1.2 Working memory

Working memory is a memory system with limited capacity for temporary processing and storage of information. Multiple theoretical perspectives regarding working memory and its limited capacity. One of the most debated is that whether working memory is unitary, domain-general process or a domain-specific process consisting of auditory working memory and visual-spatial working memory.

1.2.1 Working memory model

According to the multiple-component theory proposed by Baddeley [41], working memory could be divided into three main components: a domain-general central executive component, which is responsible for monitoring and coordinating the activities of the other two components. The phonological loop is the component responsible for the storage of auditory-based information; the visuospatial sketchpad is responsible for the storage of visual-based information. This model emphasizes that information storage occurs within specific subsystems. In addition, the episodic buffer was introduced as a new component. It is used to temporarily integrate information from different systems and link them together to form a coherent event. It allows different types of information, such as sound, images, spatial information as well as information from long-term memory, to be temporarily combined to create a meaningful event and interact with other information [41].

From the perspective of working memory processing efficiency, Daneman and

Carpenter's theoretical model proposes that the working memory capacity may be influenced by the efficiency of working memory processing. In other words, lower processing efficiency in working memory might lead to lower memory capacity [42]. For instance, when comparing expert chess players and non-expert players in a game of chess, the former often outperforms the latter, possibly due to their better working memory processing efficiency. However, if both expert and non-expert players participate in a non-chess task, this advantage tends to disappear. Overall, this model suggests that working memory lies in processing efficiency, which could be linked to the central executive system rather than the capacity for storage [43].

According to Kane and Engle's working memory model theory, working memory, distinct from short-term memory, is regarded as an executive attention mechanism [44]. Specifically, this mechanism effectively discriminates between relevant and irrelevant information, maintaining the former in a state that is easily retrievable and accessible while inhibiting the latter. Furthermore, this executive attention mechanism enables individuals to switch back and forth between different cognitive tasks, which may result in resource competition. Working memory capacity is seen as a dynamic system, not merely characterized by retention time and capacity, but related to how effectively the executive attention function can concentrate attention resources on related content. Research has provided theoretical support for this perspective, with findings indicating that individuals with higher memory capacity are more adept at directing their attention towards relevant task content while suppressing interference from irrelevant stimuli [45]. Finally, this model describes working memory in terms of executive attention.

1.2.2 Working memory with aging

The Hemispheric Asymmetry Reduction in Older Age model, or HAROLD was used to describe the brain activation during working memory task in older adults [46]. According to the HAROLD model, older adults may activate regions in the bilateral hemisphere specific to the task being performed as a compensatory mechanism during cognitive tasks. Research has found that during spatial working memory and verbal working memory tasks, young adults show lateralized activation, such as left prefrontal cortex activation (PFC) during verbal memory tasks and right prefrontal cortex activation during spatial tasks. However, older adults exhibit bilateral prefrontal cortex activation in both types of tasks. In conclusion, the PFC activity of older adults during cognitive tasks tends to exhibit reduced lateralization compared to young adults. From the perspective of compensatory mechanisms, the increased activation in specific brain regions by older adults may compensate the sensory decline during aging [47].

Moreover, the Compensation-Related Utilization of Neural Circuits Hypothesis (CRUNCH) suggests that during cognitive tasks, older adults tend to activate additional brain regions. The activation of these specific brain regions is closely associated with the performance of older individuals. It is noteworthy that this additional activation manifests as an enhancement of certain brain areas to compensate for the functional decline in other regions. For instance, increased activity in the prefrontal cortex compensates for the decline in the medial temporal lobe. However, Cappell et al. (2010) proposed that the CRUNCH is effective at lower levels of task demand. When task-demanding increases at high levels of the task, the cognitive resources may reach limitation, resulting in insufficient processing and poor performance in older adults [48]. For instance, studies utilizing working memory tasks have shown greater activation in the prefrontal and parietal cortices in older adults than young adults under relatively low working memory loads; conversely, less activity in these areas in older adults was observed under high working memory loads [49]. Furthermore, research suggests that working memory in older adults is plastic, and improvements in working memory capacity through training can lead to changes in brain structure. Specifically, working memory training can result in increased gray matter volume in older individuals. Following an 8-week memory training, the training group exhibited significant improvements in memory performance compared to the control group. At neural level, older individuals in the training group exhibited increased thickness in the right fusiform and lateral orbitofrontal cortex, which was positively correlated with the observed working memory enhancements [50]. In addition, some researchers conducted working memory training on 23 healthy older adults, with half randomly assigned to the adaptive group (high-load training) engaging in visual-spatial and verbal working memory. The other half were placed in the control group (low-load training). After five weeks, both groups exhibited improved working memory performance compared to initial level. Comparing pre-training to post- training in terms of neuroimaging results, reduced activation in specific brain regions for both groups was found, particularly pronounced in the high-load training group. Reduced activation indicates higher efficiency of brain activity, attributed to the recruitment of fewer neurons and an enhancement in their processing efficiency [51]. Moreover, other studies employed diffusion tensor imaging (DTI) techniques to investigate structural changes in older adults' brain regions resulting from working memory training. The results indicated increased DTI in the frontal lobe and enhanced connectivity of fiber bundles following training [52].

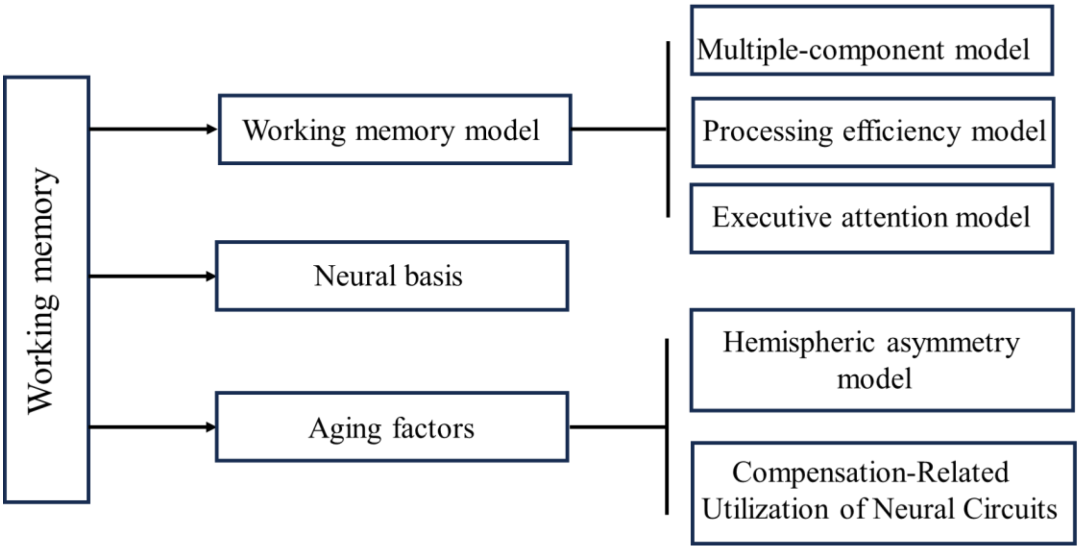


Figure 1. *An introduction of working memory consists of the following sections: working memory model, neural basis of working memory and aging factors.*

1.3 Working memory training

Neural investigations into the plasticity of working memory primarily focus on alterations in brain activation patterns. In general, working memory training can induce three distinct patterns of brain activation: (1) Training leads to enhanced brain activation, reflecting the expansion of neural structures involved in working memory processing. It is noteworthy that in older adults, the enhancement of brain region activation through working memory training may be linked to brain compensation mechanisms. (2) Training results in reduced brain activation, reflecting improved neural efficiency and enhanced processing efficiency of working memory. (3) Reorganization of brain activation, where brain areas activated at pre-training, show a combination of both enhancement and reduction in activation at post-training.

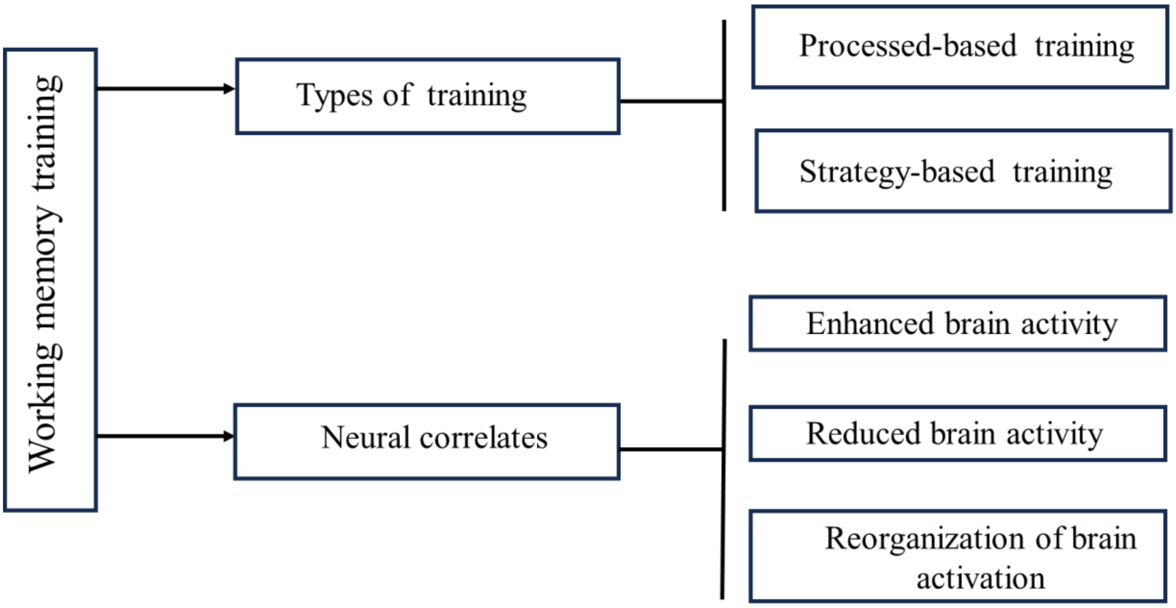


Figure 2. *An introduction of working memory training consists of the following sections: types of training, neural correlates.*

1.3.1 Neural correlates of working memory training

Numerous studies on working memory training have discovered that training leads to enhanced brain region activation, indicating higher levels of activation or the recruitment of additional cortical areas, thus facilitating cognitive processing. Some researchers have revealed that after 4-weeks of spatial working memory training, participants exhibited enhanced activation in the right inferior frontal gyrus and right intraparietal sulcus, and these activations were closely correlated with improvements in working memory [53]. Olesen et al.'s study found that following working memory training, participants exhibited increased activation in the prefrontal cortex and parietal cortex, both of which are regions associated with working memory processing [54]. Similar effects have been observed in specific populations such as ADHD patients, where working memory training led to enhanced activation in the prefrontal, parietal, and temporal regions [55].

According to the neural efficiency theory, a decrease in brain region activation may reflect the brain’s more efficient utilization of cognitive resources, resulting in resource conservation to meet the demands of other cognitive processes. Research has explored the activation changes in the striatum brought about by working memory training. As training progresses, there is an initial enhancement followed by a subsequent decline in activation in the right striatum. In contrast, the control group consistently shows increasing activation in the striatum [56]. In another neural study related to auditory n-back working memory training, a decrease in activation was observed in the right superior temporal sulcus. The authors suggested that this region is closely associated with auditory information processing, and the reduced activation resulting from training could be attributed to improved neural efficiency [57]. In another study of theirs, two groups of participants underwent visual and auditory n-back training separately. Similar patterns of brain region activation were identified, where training led to reduced activation in the right middle temporal gyrus and posterior cingulate cortex. Notably, auditory n-back training additionally resulted in decreased activation in the right middle temporal gyrus [58].

Activation reorganization was related to automatic processing and controlled processing. Specifically, training is often as a changing interplay between automatic and controlled processes. Controlled processing is governed by high-level function, such as attention and active monitoring of the task performance whereas automatic processing is activated automatically without the necessity for active control or attention [59]. When the training task is less demanding, it is easier to index a shift from controlled to more automatic processing and save limited resources for controlled processing. Such easier transition from controlled to more automatic leads to the decreased brain activity, reflecting increased neural efficiency. In contrast, when the training task was demanding, controlled processing dominates rather than automatic processing, leading to increased brain activity to ensure sufficient neural resources for controlled processing [60]. Research utilizing working memory training to investigate changes in brain functional connectivity has found that participants exhibited enhanced functional connectivity between the right middle frontal gyrus (MFG) and the anterior frontal lobe. Conversely, there was a reduction in functional connectivity between the medial PFC and the right posterior middle temporal gyrus [61].

1.3.2 Types of working memory training

Working memory training can be divided into strategy-based and process-based working memory training.

1.3.2.1 Strategy-based working memory training

Strategy-based working memory refers to enhancing working memory performance through the acquisition of strategies, which can occur during different processing stages of working memory, such as encoding or retrieval. Memory strategies include various types, such as the method of loci, imagery, association, categorization, etc. Some studies directly compared differences in working memory training between strategy-based tasks and process-based tasks. Specifically, this study utilized memory loci training and dual n-back training. The results indicated significant training effects in both groups. In transfer tasks, both groups exhibited transfer effects in the digit-span task, while only the dual n-back working memory training group showed transfer effects in the change detection task. Based on these results, the authors suggested that dual n-back training was more effective [62]. Since the study employed young participants, it might indicate that young individuals are more suited for process-based working memory training. Research also investigated age-related differences in strategy-based working memory training. Specifically, both young and older training groups received strategy training, involving enhancing word memory processing levels through learning mental imagery. The results demonstrated improved working memory performance in both training groups. Notably, the extent of improvement in working memory for the older adults was comparable to that of the young, suggesting that older adults may be suitable for strategy-based working memory training to some extent [63]. Furthermore, neural studies examined changes in brain activation due to strategy-based training. Activation in the temporal region was closely associated with strategy-based working memory training. Specifically, young adults exhibited significant activation in the temporal region following training. Moreover, the authors categorized older participants in training group based on training benefits, revealing that those older individuals who improved working memory through strategic memory exhibited enhanced activation in the temporal region, unlike the non-trained older individuals who did not show increased activation in this region. Another study investigated the influence of strategy memory training on neural activity during the memory encoding stage, finding that training enhanced brain activity in the older adults in the medial superior frontal gyrus, right precentral gyrus, and left caudate during encoding. The authors suggested that these brain regions might be related to automatic verbal encoding strategies [64].

1.3.2.2 Process-based working memory training

In contrast to strategy-based working memory training, process-based working memory training refrains from providing participants with any specific strategies. Executive functions, including switching, updating, and inhibition, can be facilitated through this type of training. Firstly, switching refers to participants switching between different tasks based on specific rules. Several studies have investigated the transfer effects of switching task training across different age groups. Specifically, the study encompassed three age ranges: 8-10 years, 18-26 years, and 62-76 years. The results revealed more pronounced near-transfer effects in both children and older adults following training. Conversely, young adults exhibited far-transfer effects, including fluid intelligence and executive control [65]. Furthermore, switching-task training was found to significantly enhance executive control functions in children with ADHD. Twenty ADHD children were randomly assigned to two training groups. One group underwent single-task training followed by switching task training, while the other group followed the opposite sequence. Both groups exhibited training effects, characterized by reduced switch costs. Transfer effects were also observed, indicating improved inhibitory control and verbal working memory levels through switching task training [66]. Their subsequent research employed a game-like format for switching task training. In this training, Task A required participants to determine whether an image depicted a fruit or a vegetable (food task), while Task B required them to judge whether an image was small or large (size task). Participants needed to switch between Tasks A and B. After 21 training sessions, elementary school students in the training group displayed training effects, which transferred to untrained executive functions. Furthermore, the reading abilities of the trained elementary school students notably improved [67]. Research also investigated brain region activation changes induced by switching task training. The results indicated reduced activation in the brain regions involved in the switching task prior to training. Only the dorsolateral prefrontal cortex exhibited increased activation. These changes in brain region activation may be associated with training effects [68].

From a working memory perspective, the primary role of inhibitory function is to suppress irrelevant information from entering the working memory processing. Firstly, research has investigated the training effects of inhibition on both young adults and children. Inhibition training employed three Stroop tasks. The stimuli for these tasks consisted of landscape images (associated with the “day” response key “F” and the “night” response key “J”), fruit images (associated with the “big” response key “S” and the “small” response key “L”), and cartoon images (associated with the “boy” response key “D” and the “girl” response key “K”). Participants were required to respond to the opposite category of the currently presented image, for instance, pressing the "D"key when a girl image was presented, corresponding to "boy". Results showed that both training groups exhibited wide-ranging transfer effects following the training. The authors also examined the durability of these transfer effects and found that all of these transfer effects dissipated over time [69].

The updating function involves continuously modifying the contents of working memory based on newly presented information. It primarily monitors incoming information, replacing old information in working memory that is less relevant to the current task with new information that is highly task-relevant, thus constantly revising memory content [70]. The dual n-back, single n-back are widely employed as measurements of working memory updating. The dual n-back task requires participants to process stimuli streams from both the visual and auditory modalities simultaneously and determine whether the current stimuli in both modalities match the stimuli presented N trials ago. Importantly, many studies employ an adaptive training approach, in which the level of difficulty increases as participants achieve higher accuracy rates. For instance, if a participant performs well at a 2-back level, the difficulty level is up to 3-back. A classic dual n-back training paradigm comes from Jaeggi and colleagues. In this task, sequences of squares are presented in different positions on the screen, accompanied by the auditory presentation of different letters' sounds. Participants are required to simultaneously remember the location of the squares they see and the letters they hear and compare the current square and letter with the one presented N trials ago [71]. A key press response is required if they match. N adapts based on the participant's memory performance, increasing, or decreasing as memory performance improves or deteriorates. Larger N values require participants to hold more information in working memory, thereby increasing task difficulty. Using this training task, the authors investigated the impact of working memory training on fluid intelligence. The training group was divided into four subgroups based on training duration: 8 days, 12 days, 17 days, and 19 days. They found that, following working memory training, the training groups (excluding the 8-day training) exhibited significant improvements in fluid intelligence scores compared to the control group that did not receive training. Interestingly, the longer the training duration, the greater the transfer effect on fluid intelligence for the training group. However, some studies have presented opposing conclusions. In a 5-week dual n-back training, it was found that the training group (30-60 years) did not exhibit significant transfer effects on fluid intelligence[71]. In another study investigating differences in training effects between older adults and young adults on the dual n-back task, it was found that young adults benefited significantly more from dual n-back training compared to older adults. Notably, significant transfer effects were not observed in older adults for task switching and inhibition tasks [72]. This could be attributed to the fact that the dual n-back task continually demands the allocation of cognitive resources to expand cognitive functions, approaching cognitive capacity limits, thus presenting higher task difficulty. Furthermore, some studies directly compared the differences in training and transfer effects between dual n-back training and visual n-back training. The results revealed similar training effects for both tasks, and even better transfer effects to the updating task were observed after visual n-back training [73].

Some researchers have employed ERPs to investigate neural changes induced by N-back training. In a study by Covey (2018), N-back training was contrasted with visual search training to explore the neural mechanisms underlying working memory training based on the N-back task. Following 4 weeks of training, both the N-back training group and the visual search training group exhibited significant enhancements in P300 amplitude. Additionally, the N-back training group displayed increased N200 amplitude and an earlier latency. The authors suggested that the enhanced P300 amplitude in both training groups might be linked to improved attention. The alteration in the N200 component within the N-back training group could be attributed to the specific nature of the N-back task, where the N200 reflects enhanced conflict detection and mismatch identification abilities [74]. Regarding the neural differences between older adults and young adults, a study employed N-back training for investigation. The participants were divided into older training groups, young training groups, older control groups, and young control groups. The results revealed significant post-training enhancement in P300 amplitude for both older adults and young adults. Both training groups exhibited near-transfer and far-transfer effects on working memory and fluid intelligence, with greater training benefits observed in older adults for attention shifting tasks [75]. Other studies have explored neural plasticity in children. One study recruited 44 children to participate in digit-span backward and N-back training. The N-back task was used to record EEG data before and after training, to assess neural changes brought about by training. After 20 training sessions, the children exhibited a noticeable increase in P300 amplitude during the 2-back task. Furthermore, studies have investigated changes in brain region activation induced by N-back training tasks [76]. Due to the fact that n-back task was associated with updating, fMRI studies have localized the updating function in the striatal area [77]. Another study confirmed this idea, using the dual n-back training task. It found that the training group exhibited training effects on the untrained bimodal updating task, but not on the untrained unimodal updating task. Additionally, for the training group, training led to increased striatal activation, which was strongly correlated with improvements in working memory [78]. Single n-back training has also been used to investigate changes in brain region activation. Compared to the control group, the N-back training group displayed significant improvements in working memory performance, reflected in increased accuracy and reduced reaction times. Moreover, the training group exhibited reduced activation in frontal superior/middle cortex, inferior parietal cortex, anterior cingulate cortex, and middle temporal cortex. The authors proposed a close relationship between behavioral improvements and the reduced activation in these brain regions. Additionally, the study suggested that this training effect was well-maintained, as a follow-up fMRI scan conducted 5 weeks later still showed stable behavioral and brain activation changes [79]. Another study investigated the changes in brain region activation brought about by N-back training across three-time stages: pre-training, mid-training, and post-training. After undergoing 10 sessions of working memory training, it was concluded that the training effects of the N-back task are closely linked to the activation of the parietal region responsible for working memory updating. It's worth noting that certain studies incorporate multiple components of working memory, such as updating and inhibition components, into training. For instance, a study employed both the flanker training task and the N-back training task to enhance individuals' interference control abilities. The training was divided into an adaptive group and a non-adaptive group. The results revealed that adaptive training led to improved interference control abilities, reflected by reduced error rates and reaction times in the flanker task. On a neural level, adaptive training resulted in a significant increase in N200 amplitude for incongruent trials in the flanker task, along with a decrease in CRN amplitude. Conversely, the non-adaptive group exhibited enhanced reaction speeds but also a rise in error rates [80].

1.4 Audiovisual integration and working memory

Before understanding the relationship between audiovisual integration and working memory training, the common between audiovisual integration and the working memory process should be taken into consideration. Previous research on audiovisual integration has found that individuals with larger working memory capacity can achieve greater audiovisual integration. They can better rely on visual and auditory cues to perceive speech, even in situations with higher auditory noise levels. This indicates that working memory capacity plays a regulatory role in individual audiovisual integration abilities: higher working memory levels can provide a certain inhibitory effect on noise, allowing individuals to make better use of audiovisual cues [81]. Other studies have employed simultaneity judgment tasks to explore the association between audiovisual processing and working memory. In this task, participants are required to judge whether presented visual and auditory stimuli are synchronous and compute their simultaneity perception sensitivity. Additionally, errors in responses are recorded for each stimulus onset asynchrony (SOA) condition. This research suggests a close relationship between simultaneity perception sensitivity and multisensory processing, while errors in responses for each SOA condition are closely related to executive control within working memory. Results reveal a significant correlation between increased error responses and lower working memory capacity in middle-aged and older individuals. This implies that there might be shared cognitive resources between audiovisual simultaneity judgment and working memory processing [82].

Moreover, working memory also exhibits characteristics of audiovisual integration. For instance, a recognition task was conducted to examine how the simultaneous presentation of semantically congruent audiovisual information (e.g., presenting a picture along with its corresponding sound) compared to visually and auditorily congruent non-semantic information affects memory recognition. The results showed that memory recognition was better when semantically congruent audiovisual information was presented compared to when audiovisual congruent non-semantic information was presented. Specifically, the congruency of semantic audiovisual information facilitated the encoding process, leading to improved memory performance in children. They proposed that semantically congruent audiovisual information helps children establish associations between visual and auditory information, facilitating rapid linkage between perceptual information and long-term memory and thereby enhancing recognition [83]. Research by Postle and others has indicated that brain regions involved in perceptual processing also play a role in storing that perceptual information [84]. For instance, areas responsible for visual processing might also serve as short-term memory storage regions for visual information. In tasks related to facial information memory, activation of the posterior fusiform gyrus has been observed [85], which is closely associated with facial recognition processing. To further investigate whether working memory representation is amodal or modality-specific, Xie and colleagues employed a Sternberg-type working memory task to examine changes in brain activation during object recognition under conditions of semantic congruence and incongruence. The Sternberg task involved three working memory processing stages. In the cue encoding stage, a cue is presented to indicate the upcoming type of stimulus, followed by a 2-second encoding stage where participants must remember the presented stimulus. During the delay stage, participants retain the encoded information. Finally, a retrieval stage involves determining whether the presented stimulus matches the stimulus from the encoding stage. After comparing brain activations across four conditions (audiovisual congruent, audiovisual incongruent, visual only, auditory only), it was found that under audiovisual congruent encoding, shared activated regions included the left angular gyrus, supramarginal gyrus, and precuneus. Under audiovisual incongruent encoding, shared activated regions included the bilateral angular, left superior parietal lobule, and left middle temporal gyrus. Further examination of commonly activated brain regions under congruent and incongruent conditions revealed bilateral angular gyrus and left middle frontal gyrus n in both conditions. This study supported amodal representation of working memory [86]. Moreover, early ERPs findings also provided evidence for the facilitative effect of audiovisual integration on working memory. A study investigated whether audiovisual integration could enhance working memory performance during aging. Participants were required to complete an N-back task in visual, auditory, and audiovisual conditions. In the visual condition, only visual digits were presented; in the auditory condition, only auditory digit sounds were presented; in the audiovisual condition, participants were presented with both visual digits and corresponding sounds. The results revealed significantly higher working memory performance in the audiovisual condition compared to the visual and auditory conditions. The correct response rate for working memory in older adults under the audiovisual condition was relatively comparable to that of young adults under the auditory condition, suggesting that the presentation of audiovisual stimuli can help older adults mitigate age-related working memory decline. Further neural results found that the N1 amplitude for older adults in the audiovisual condition was significantly lower than the N1 amplitude in the auditory condition, indicating the presence of the audiovisual facilitation effect at the neural level. Additionally, both older adults and young adults exhibited an earlier N1 latency in the audiovisual condition compared to the auditory condition, suggesting to some extent the facilitative effect of audiovisual integration [87].

Recently, several studies have employed the resting-state homogeneity method to investigate the neural mechanisms of audiovisual working memory. In this research, the Sternberg working memory task was employed with four different encoding conditions:

audiovisual congruent, audiovisual incongruent, visual only, and auditory only conditions. The results continued to demonstrate the facilitative role of audiovisual integration on working memory. Specifically, presenting audiovisual stimuli during the encoding stage enhanced subsequent working memory maintenance and retrieval. The study further revealed that in the high-performance working memory group, the REHO (Regional Homogeneity) values of the executive control network significantly increased, while those of the default mode network and salience network significantly decreased. Thus, the authors hypothesized that the facilitative effect of audiovisual integration on working memory may involve activation of different functional networks, such as the executive control network [88]. Another study employed source localization methods to investigate how the presentation of audiovisual stimuli during the encoding stage impacts subsequent working memory processing. This experiment, based on the Sternberg paradigm, also found enhanced working memory performance under the audiovisual congruent condition. Furthermore, neural results revealed significant differences in the ERPs between the audiovisual congruent and the sum of visual and auditory conditions in the frontal and parieto-occipital regions at a relatively late 236-530 ms time window. This difference was likely associated with the audiovisual effect. The authors further explained the reason for the appearance of this difference in the late stage, citing earlier research on audiovisual integration that indicated its occurrence before 100 ms, which may be associated with reflecting bottom-up stimulus characteristics. In contrast, this study found that the audiovisual difference was at a relatively late stage, suggesting top-down processing involvement [88]. In their subsequent neural oscillation study, the researchers found that during audiovisual encoding, theta neural oscillations appeared in multiple brain regions, including the frontal lobe (superior frontal gyrus), parietal lobe (precuneus), temporal lobe (inferior temporal gyrus), and occipital lobe (cuneus). This highlighted the vital role of theta oscillations in encoding under audiovisual conditions and in aiding the formation of audiovisual memory traces [89]. However, other studies have discovered that working memory load can modulate neural oscillations associated with audiovisual integration. Specifically, in a dual-task experiment involving an N-back task and sound-induced flash illusion, it was investigated whether working memory load affects audiovisual processing (top-down or bottom-up). The results showed that under high working memory load, not only did more illusion effects occur, but they also influenced early and late neural oscillations. Particularly, under high working memory load conditions, early beta neural oscillations in the auditory and motor area were found at 70 ms, and the theta neural oscillation was increased at 120 ms, and late beta neural oscillations in the frontal and auditory cortices during the illusion were suppressed. The authors suggested that audiovisual integration can be influenced by cognitive resource availability. When cognitive resources are limited, audiovisual integration may involve top-down neural oscillations including beta/theta [90]. Similar audiovisual effect has also been identified in working memory. Researchers used a combined n-back and Go/NoGo paradigm to investigate whether audiovisual interaction interferes with or facilitates working memory. In detail, the experiment had two groups: one completed a single visual n-back task and a single auditory go/no-go task, while the other completed a dual task combining N-back and Go/NoGo. Comparing single visual 2-back with the dual task revealed no difference in accuracy but differences in reaction time. The authors found a significant correlation between reaction time and accuracy, suggesting a speed-accuracy trade-off in the dual task. There was no significant difference in visual working memory performance between the two tasks at 2-back. It appears that auditory did not interfere with visual. However, comparing single visual 3-back with the dual task showed a significant decrease in accuracy and reaction time in the dual task. Additionally, in the dual task, visual memory significantly declined. This suggested that under relatively high load (3-back), auditory stimuli began to interfere with visual stimuli. The authors proposed that under high load, the brain couldn't simultaneously process visual and auditory stimuli due to limited cognitive resources, leading to decreased performance [91].

Based on audiovisual effect on working memory, whether audiovisual working memory could induce a training effect or a transfer effect? Some studies have provided neural explanations for the brain activation resulting from multisensory training. In one study, participants underwent fMRI scans during three processing stages: the pre-associative learning stage, the associative learning stage, and the post-associative learning stage. During the associative learning stage, participants engaged in audiovisual associative learning tasks, such as associating sounds with faces or phone rings with phone brand names. The results revealed that participants who underwent audiovisual associative learning showed improved performance in recognizing sounds associated with faces, along with enhanced functional connectivity between auditory and visual regions. The authors proposed that the presentation of audiovisual stimuli (sounds and faces) triggers multisensory associative effects. This might be driven by predictive coding, where the activity of higher-level neurons influences early perceptual processing through feedback loops. In other words, the positive impact of higher-level neural activity from audiovisual associative learning might benefit lower-level visual recognition [92]. Why does multisensory training lead to better facilitative effects? First, it's argued that perceptual information is encoded in specific brain regions closely linked to multisensory integration areas. When training involves audiovisual stimuli, more multisensory regions are activated compared to unimodal training. This results in stronger memory traces, which subsequently influence later processing, creating facilitative effects. However, whether training based on audiovisual benefits can enhance higher-level cognitive processes, such as working memory, remains largely unexplored. In a study by Deveau et al. (2016), the authors suggested that the design of working memory training should follow multisensory principles. They noted that memory storage and retrieval involve multiple senses. They argued that existing working memory training tasks like the dual n-back don't effectively harness the potential for learning facilitation. This is because such tasks require visual and auditory processing to compete for limited resources, leading to substantial interference effects. They proposed that multisensory training can tap into complementary information from different modalities. For instance, individuals with limited visual abilities could benefit from training involving associated auditory stimuli [93].

1.5 The aims of the present thesis

The main aims of the present thesis can be divided into three sections:

Experiment 1 introduces the concept of audiovisual integration and the relationship between audiovisual integration and age-related factors. We also test whether audiovisual integration could occur in young adult and older adults, measured by behavioral results and ERPs. Moreover, we also portrayed the different magnitude of audiovisual integration between young adults and older adults.

Experiment 2 introduces the benefit of audiovisual working memory. Given the fact that whether working memory with audiovisual stimuli could enhance working memory performance remained in conflict, we test the behavioral difference between audiovisual working memory and single-modality working memory (auditory, visual) in older adults and young adults.

Experiment 3 introduces the audiovisual benefit on working memory training. We test whether working memory training with audiovisual stimuli could improve the working memory performance in young adults and older adults. Moreover, we also test whether the training of higher-cognitive processing could transfer to low-cognitive processing. That is, the transfer effect of working memory training on audiovisual processing.

Chapter 2 The alteration of audiovisual integration in young and older adults

2.1 Background

In our daily life, we receive information from different sensory modalities. Our brain has the capacity of integrating these information sources into a unique and coherent percept, this process of information integrating is called multisensory integration, a phenomenon that shapes our view of the world [94]. For example, a train rumbles quickly across the plain which conveys the visual and auditory information, but we recognize the train by integrating the information of different sensory modalities instead of perceiving the train as fragmented shapes, colors, and sounds.

The stimuli intensity can be associated with multisensory integration processing. Many behavioral studies for young people of multisensory integration have also shown that the employment of bimodal stimuli with low intensities produced more stronger behavioral enhancement than the employment of bimodal stimuli with high intensities [103,104]. The Followed ERPs research referring to the neurophysiologic basis for the relationship between stimulus intensity and the magnitude of the multisensory integration revealed that audiovisual integration was found particularly for stimuli with low intensity, and high intensity level did not reveal robust audiovisual integration [105]. Some researchers suggested such phenomenon that enhanced audiovisual integration occurred in low intensity might be due to the inverse effectiveness, where multisensory integration was most effective and yielded maximal behavioral enhancements when less intense or weak and ambiguous stimuli were employed [106]. However, other findings related to aged group found that the inverse effectiveness was not observed when the stimuli intensity decreased [107,108]. Particularly, the study of Tye-Murray et al. (2010) compared the ability of young adults and older adults to integrate auditory and visual sentence materials under conditions of good and poor signal clarity, observing whether the Inverse Effectiveness occurred or not. Neither integration enhancement increased when signal clarity in the auditory or visual channel of audiovisual speech stimuli decreased, nor was either higher for older adults than for young adults [107]. This research figured out the results in old group was due in large part to the test format or context, rather than the further investigation that how stimuli intensity influences the audiovisual integration during aging. Because, in the context of real-world multisensory stimuli, the changes in multisensory integration can be mediated not only by changes in the external characteristics of the stimuli (e.g., the loudness of the auditory or the clarity of the visual stimulus) but also by changes in internal events governing the processing of the information - the decreases in visual and auditory sensitivity associated with aging likely stem from reductions in internal signal strength caused by alterations in transduction and encoding mechanisms [27].

As we age, our life span undergoes a decline of function in our sense, such as impairment of visual acuity [95] and the increase of auditory thresholds [96], which may influence the effectiveness of audiovisual integration and processing. This phenomenon is mainly manifested as enhanced or reduced audiovisual integration in older adults compared to young adults. Specifically, in a magnetoencephalography (MEG) study, brain activation differences in audiovisual processing between older and young adults were measured in the auditory cortex [97]. At behavioral level, older adults exhibited slower overall response times than young adults. When assessing differences in audiovisual integration using the race-model, it was found that the race-model suggested that audiovisual integration is reduced in older adults compared to young adults. In further investigation of neural results, it was discovered that older adults exhibited cortical suppression when responding to audiovisual stimuli, as evidenced by significantly reduced amplitudes compared to young adults. The authors further proposed that this cortical suppression contributes to the weakened audiovisual integration ability observed in older adults [97]. Nevertheless, numerous other studies have indicated that older adults might have better audiovisual integration abilities compared to young adults. For instance, audiovisual discrimination, sound-induced flash illusion tasks, and speech perception tasks have demonstrated this enhanced audiovisual integration in older adults. Some researchers employed a colour-naming task with semantic stimuli presented under visual, auditory, and audiovisual conditions Visual stimuli consisted of discs (red, blue, green), auditory stimuli were corresponding sounds, and audiovisual stimuli combined both visual and auditory stimuli. The findings showed reduced reaction times for audiovisual stimuli in both young and older adults [28]. In a study related to audiovisual integration using hand-held tools as stimuli, participants were required to discriminate between target stimuli (hand-held hammer) and irrelevant standard stimuli (hand-held stick), presented through visual, auditory, and audiovisual modalities. At neural level, they found that greater audiovisual integration of anterior brain was found in older adults in the low-intensity condition. This suggests that audiovisual integration in older adults works despite reduced clarity. Furthermore, using race-model, greater audiovisual integration benefits for older adults was found compared to young adults [98]. In a study examining age-related differences in the McGurk effect, the accuracy of phoneme recognition was investigated under audiovisual congruent and incongruent conditions. The experiment included congruent audiovisual, incongruent audiovisual, single visual, and single auditory conditions. The results suggested that older individuals appeared more susceptible to the influence of visual speech, leading to larger illusions and greater audiovisual integration. This enhancement in audiovisual integration among older individuals may be attributed to the following reasons: firstly, it could serve as compensation for age-related sensory decline [27]. Secondly, the enhanced audiovisual integration might stem from a wider binding window in older individuals. This window refers to the time interval within which the central nervous system optimally integrates information for better perception of external events. When visual and auditory information is presented within a certain time window, integration occurs [9]. Older adults might have a faster temporal binding window, leading to a higher likelihood of integration and consequently, the generation of illusions. In the aforementioned studies, the enhanced audiovisual integration in older adults may be related to the meaningful stimuli these studies used. Some researchers have particularly investigated the impact of stimulus types on audiovisual integration in older adults. The study investigated the differences in audiovisual integration between older and young individuals under conditions of audiovisual meaningless stimulus presented at the center and audiovisual semantic stimulus presented at the center). The research findings indicated that the response times to audiovisual targets were significantly shorter than those to auditory or visual targets. Young participants exhibited significantly shorter target response times across all modalities compared to older participants. Furthermore, the results from the race model revealed that older participants exhibited delayed and broader audiovisual integration windows in all conditions compared to young participants. Lastly, within the context of simple meaningless stimuli, older adults exhibited the smaller audiovisual integration, whereas in the semantic stimuli, older adults displayed the greater audiovisual integration. This suggests that despite the compromised audiovisual integration ability in older adults, a compensatory mechanism is established when processing audiovisual semantic stimuli. The authors suggested that older adults engage in audiovisual integration processing involving meaningful stimuli, they activate more brain regions to facilitate behavioral performance, indicating their ability to utilize knowledge and experience to integrate stimulus information from different modality [30]. However, the study did not utilize neural techniques to further investigate the changes in audiovisual integration in older adults when using simple stimuli. This raises a question: do older adults activate more brain regions and exhibit greater audiovisual integration in processing meaningless audiovisual stimuli using ERPs?

In order to clarify the aging effect on audiovisual integration between older adults and young adults when processing simple meaningless stimuli, it was hypothesized that old group would have more gains in audiovisual integration compared to the young group in the audiovisual condition predicted by the idea that multisensory stimuli may offer a performance benefit due to the deleterious sensory systems. Here, we conducted an ERP study designed to demonstrate how age affects audiovisual integration in meaningless stimuli. The participants performed a simple discrimination task based on audiovisual, auditory, and visual stimuli conditions. This study may help resolve the disputes on audiovisual integration mentioned in previous studies.

2.2 Methods

2.2.1 Participants

Twenty-four old (age ranging from 58 to 70, mean age ± SD, 61 ± 4) and 21 young participants (age ranging from 20 to 25, mean age ± SD, 21.7 ± 2.5) were recruited to take part in the study. One participant was excluded from the analysis due to poor performance in ERP data collecting (the loss of > 70% of trials). The remaining 43 participants had normal hearing and normal or corrected-to-normal vision. All participants were naive to the purposes of the study and provided informed consent for the procedure, which was previously approved by the Ethics Committee of Hubei University.

2.2.2 Stimuli

We conducted a pre-experiment prior to all experiments to determine one contrast level whose accuracy in detecting the visual stimuli ranges from 80% to 90%. All the stimuli were divided into two types—target stimuli and standard stimuli. The horizontal gabor pitch and the vertical pitch were regarded as target stimulus (VT) or standard stimulus (VS) respectively, which appeared 5° below the fixation point. The auditory target stimuli (AT) were white noise and the auditory standard stimulus (AS) was a 1000 Hz sinusoidal tone. These stimuli were presented simultaneously in both the ears through an earphone. The audiovisual standard stimulus (AVS) consisted of the simultaneous presentation of unimodal A and V standard stimuli. And the audiovisual target stimulus (AVT) consisted of the simultaneous presentation of unimodal A and V target stimuli.

2.2.3 Procedure

The experiment was conducted in a dark and sound-attenuated room. There were 200 visual stimuli (40VT, 160VS), 200 auditory stimuli (40AT, 160AS) and 200 audiovisual stimuli (40AVT, 160AVS). The stimulus stream consisting of visual stimuli, auditory stimuli, and audiovisual stimuli, which were segmented into 3 sessions, was randomly presented on monitor using Presentation software (Neurobehavioral Systems Inc., Albany, California, USA). AS shown in Figure 1, Each trial began with the presentation of a fixation cross at the center of the screen, accompanying a type of stimulus whose duration was 400 ms. The inter stimulus interval ranged between 1200 and 1800 ms. The participants, viewing the monitor from 60 cm, fixing their eyes on the fixation point, were asked to response to target stimuli as quickly and accurately as possible by pressing the left button of mouse.

The EEG and behavioral data were recorded simultaneously. An EEG was captured through a 32-channel cap (Easy-cap, Herrsching Breitbrunn, Germany) using a BrainAmp MR PLUS EEG Amplifier (Gilching, Germany). All electrode impedance was maintained below 5 kΩ. The electrooculogram was recorded by placing one electrode at the outer canthus of the left eye to monitor horizontal eye movements. Additionally, eye blinks and vertical eye movements were tracked using an electrode positioned beneath the right eye. Raw signals were digitized using a sample frequency of 500 Hz. The ERPs analysis was based on Brain Vision Analyzer software (version 1.05, Brain Products GmbH, Munich, Bavaria, Germany).

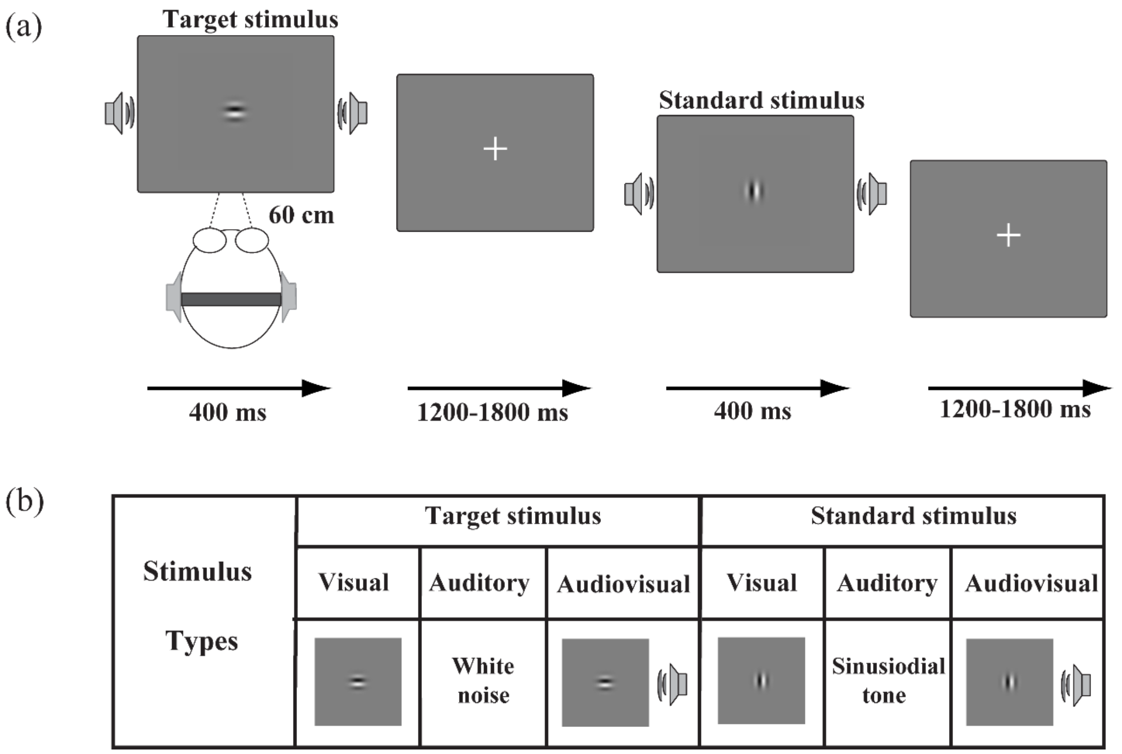


Figure 1. *The participants, viewing the monitor from 60 cm, fixing their eyes on the fixation point, responded to VT stimuli, AT stimuli or AVT stimuli.*

2.2.4 Data analysis

2.2.4.1 Behavioral data

Reaction times (RTs) and hit rates (HRs) for target stimulus were calculated for each type of stimulus and each age group. The behavioral results of RTs and HRs were used to do further analysis according to a repeated-measures analysis of variance, comprising of within-factor (three types: visual, auditory, audiovisual), and the between-factor (two group: young, old).

The race model was used to determine if the observed multisensory behavioral enhancements were beyond that predicted by statistical summation for the unimodal visual and auditory conditions. The cumulative distribution functions (CDFs) were employed to calculate the cumulative distribution of different modalities. Furthermore, for each participant, we compare the AVT CDFs to CDFs of summation for VT and AT at each time bin, using the quantitative formula of race model : [P (AT) × P (VT)] - [P (AT) + P (VT)] [99], P(AT) represents the probability of reaction time to each AT, and P(VT) represents the probability of reaction time to each VT. If the probability of reaction time to AVT was significantly greater than that predicted by the summed probability of VT and AT, suggesting that audiovisual integration occurred.

Moreover, the probability of deviations was obtained by subtracting the probability of summation for VT and AT from AVT probability at each time bin for each participant, and were averaged at each time bin, generating a difference curve for each age group. Then one-sample t test was conducted at each time bin within each age group to reveal the time range in which the curve was significantly greater than zero. Based on the probability of deviations, the positive area under the curve (AUC) can be calculated to further compare the effect of audiovisual integration in different age group or different intensity.

2.2.4.2 Behavioral data

The standard stimuli were used to elicit the ERPs to do further analysis. All signals were referenced to FCz. All electrodes were referenced to the average of both mastoids using offline analysis. A digital bandpass filter was employed with a range of 0.01-60 Hz. The data were divided into epochs, from 100 ms before the stimulus onset to 800 ms after the stimulus onset, and baseline corrections were made from −100 to 0 ms relative to stimulus onset. The artificial correction was set with a range of ±100 μV. Then the data were averaged for each stimulus type. In addition, after filtering the data with a bandpass filter of 0.1-30 Hz, the participant who lost 70% trails of one type stimulus were excluded. The grand-averaged data were obtained for each stimulus type in each electrode of each age group.

The audiovisual integration effect was measured by the formula [AV − (A + V)]. The amplitudes were derived for each time point and electrode. Firstly, pairwise comparisons using paired t-tests were conducted among all the time points for each electrode. According to previous study, significant differences can be regarded as differences in ≥ 24 ms consecutive time points from zero (at least 12 data points) [100]. The time range and the regions of interests (ROIs) we chose depends on the results of t-test in which reveals the significant audiovisual integration. Five regions of Interest (frontal: F7, F3, Fz, F4, F8; fronto-central: FC5, FC1, FC2, FC6, central: C3, Cz, C4 and occipital: O1, Oz, O2, centro-parietal: CP5, CP1, CP2, CP6) were selected based on statistical analysis. The young and older groups were analyzed separately, graphing the result of the P-value of a paired t test between AV and (A + V) at each electrode site in each group. Secondly, repeated measures ANOVAs were conducted separately for each age group at each condition, which were selected based on an overview of the significant differences in the first step. The mean amplitude data were analyzed with factors of stimulus type (AV, A+V) and ROIs. If a significant interaction between stimulus type and ROIs was observed for the main time intervals, the ANOVAs were measured separately for each of the five ROIs using the factor stimulus type (AV, A+V).

2.3 Results

2.3.1 Hit rates

For both the young and old participants, hit rates were greater than 85% for the auditory, visual and multisensory conditions (see Table 1). A repeated-measures 2 age (young, old) × 3 modality (visual, auditory, audiovisual) on hit rates revealed significant main effect of modality, $F(2, 82) = 5.23$, $p < 0.01$, $\eta^2 = 0.113$. There was no main effect of age group, $F(1, 41) = 2.63$, $p > 0.05$, $\eta^2 = 0.06$. and no significant interaction between age group and modality, $F(2, 82) = 0.775$, $p > 0.05$, $\eta^2 = 0.019$.

2.3.2 Response times

The mean response times for all conditions are presented in Table 1. A repeated-measures analysis 2 age (young, old) × 3 modality (visual, auditory, audiovisual) on the mean response times (RTs) revealed significant main effect of age group, $F(1, 41)= 12.18$, $p < 0.001$, $\eta^2 = 0.229$, indicating that the response time of the old participants was significantly slower than that of the young participants. A significant main effect of modality occurred, $F(2, 82) = 120.44$, $p < 0.001$, $\eta^2 = 0.746$, indicating that the response time declined with the alteration of modality, an alteration from unimodal modality to bimodal modality. However, the interaction of age group and modality was not significant $F(2, 82) = 1.59$, $p > 0.05$, $\eta^2 = 0.037$.

Table 1 *The mean response time and hit rates for both young and old group*

	Auditory	Visual	Audiovisual
Older group			
Hit rate (%)	87 (22)	91 (14)	97 (1.4)
RT (ms)	635 (79.1)	661 (63.4)	525 (50.8)
Young group			
Hit rate (%)	94 (6.9)	94 (6.8)	99 (1.5)
RT (ms)	557 (61.8)	598 (73.4)	477 (51.3)

Standard deviations are given in parentheses

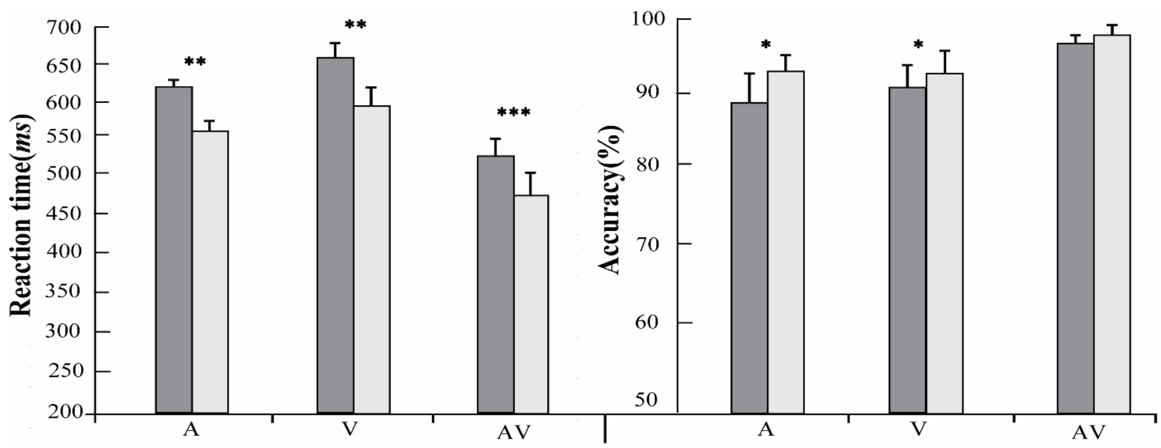


Figure 2. *Reaction time and accuracy across auditory, visual and audiovisual conditions in both older group and young group. The dark grey square represented the older group while the light grey square represented the young group.*

2.3.3 Race model

Comparing the audiovisual RTs distribution to the summation for the unisensory visual and auditory RTs distribution, significant difference was found from 370-670 ms in young group and from 380-740 ms in old group. Two-sample t test was performed to compare both age group. The result showed that there was a significant difference between the young group with the peak latency 17.98% at 470 ms and the old group with the peak latency 26% at 520 ms, $t(20) = 2.25$, $p < 0.01$. Then Independent-samples t test of AUC showed that audiovisual integration enhancement was stronger in old group (Auc = 13, SD = 5.17) than that in young group (Auc = 9.2, SD = 6.8), $t(40) = 2.03$, $p < 0.05$.

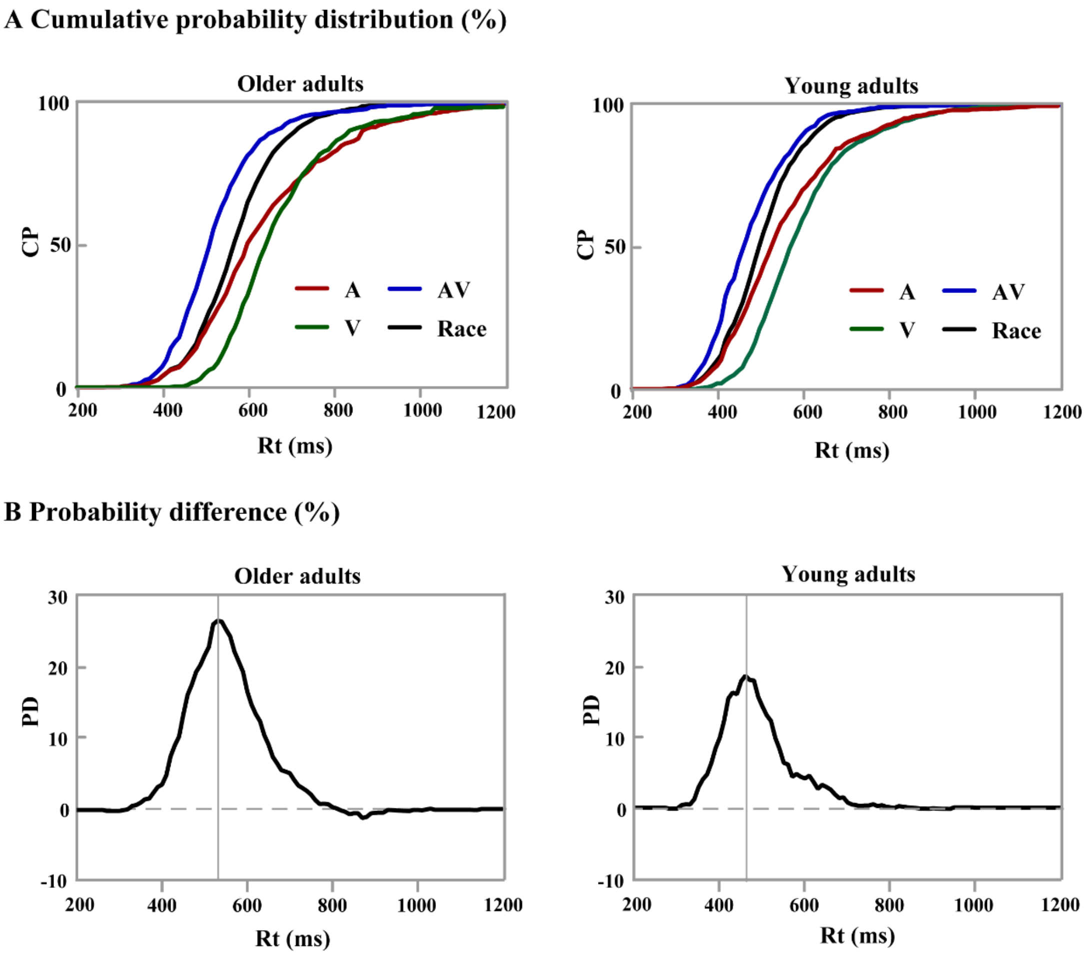


Figure 3. *Cumulative probability distribution in older and young adults (A) and Probability difference in older and young adults (B).*

2.3.4 ERP Results

The audiovisual integration was measured by [AV − (A + V)]. If the amplitudes of [AV − (A + V)] were significantly different from zero in the electrodes at 24 ms consecutive time points or more, the audiovisual integration occurred. Specifically, the audiovisual integration in old group were occurred at three-time intervals: 110-140 ms, 140-170 ms, 350-380 ms while these in young group were observed at two-time intervals: 190-250 ms, 300-360 ms.

For older adults, at the audiovisual integration for 110-140 ms, the observed integration was in the frontal (F3, Fz, F4, F8), fronto-central (FC1, FC2, FC6, FC5), central (C3, Cz, C4), centro-parietal (CP1, CP2, CP6) regions. The ANOVA using the factors stimulus type (AV and A+V) and the ROIs (frontal, fronto-central, central, centro-parietal) showed a significant main effect of factors stimulus, $F(1, 66) = 26.9$, $p < 0.001$, $ŋ^2 = 0.551$, suggesting that the amplitude of AV was significant greater than that of (A+V). And a significant main effect of ROI regions, $F(3, 66) = 128.21$, $p < 0.001$, $ŋ^2 = 0.765$, revealing that centro-parietal region had the strongest difference of amplitudes compared to other three regions. There was also a significant interaction between two factors, $F(3, 66) = 4.7$, $p < 0.01$, $ŋ^2 = 0.176$. Further post-hoc comparisons showed that amplitudes of the frontal, fronto-central, the central and centro-parietal (all p < 0.001) site were significantly greater in AV (mean potential: -4.06 µV, -3.11 µV, -2.41 µV, -1.33 µV) compared with (A+V) (mean potential: -7.4 µV, -6.68 µV, -5.06 µV, -4.11 µV).

The audiovisual integration was also found at 140-170 ms, the observed integration was in the central (C3, C4), centro-parietal (CP5, CP1, CP2, CP6) regions. The ANOVA analysis 2 (stimulus type: AV, A+V) × 2 (RIO regions: central, centro-parietal) revealed a significant main effect of stimulus type, $F(1, 22) = 38.7$, $p < 0.001$,

$\eta^2 = 0.638$, indicating that the amplitudes of AV were significant positive than (A+V) stimulus. There was also a significant main effect of RIO regions $F(1, 22) = 302.5$, $p < 0.001$, $\eta^2 = 0.932$, which demonstrated that centro-parietal region had stronger amplitude than central region. The interaction between the two factors was non-significant, $F(1, 22)= 3.1$, $p > 0.05$, $\eta^2 = 0.125$.

Referring to the audiovisual integration for 350-380 ms, the observed integration were in the frontal (F3, Fz, F4, F7), fronto-central (FC1, FC2, FC5), the central (C3, C4, Cz), centro-parietal (CP5, CP1, CP2, CP6), the ANOVA analysis 2 (stimulus type: AV, A+V) × 4 (RIO regions: frontal, central, fronto-central, centro-parietal) revealed a significant main effect of stimulus type, $F(1, 66) = 38.1$, $p < 0.001$, $\eta^2 = 0.505$, indicating that the amplitudes of AV were significant positive than (A+V) stimulus. There was also a significant main effect of RIO regions, $F(3, 66) = 123.5$, $p < 0.001$, $\eta^2 = 0.704$, which demonstrated that centro-parietal region had the strongest amplitudes compared to other three regions. The interaction between the two factors was also significant, $F(3, 66) = 8.6$, $p < 0.001$, $\eta^2 = 0.282$. The post-hoc comparisons were conducted to reveal the variance of different amplitudes at different regions. The amplitudes of frontal ($p < 0.05$), central ($p < 0.001$), fronto-central ($p < 0.001$), centro-parietal ($p < 0.001$) site were more positive in AV (mean potential: -2.93 μV, -1.83 μV, -0.74 μV, 0.76 μV) compared with (A+V) (mean potential: -4.62 μV, -4.58 μV, -3.4 μV, -0.87 μV).

For young adults, at time interval 190-250 ms, the observed integration were in the frontal (F3, Fz, F4), fronto-central (FC1, FC2, FC5), the central (C3, C4, Cz), centro-parietal (CP5, CP1, CP2, CP6), occipital (O1, Oz) regions, a ANOVA analysis 2 (stimulus type: AV, A+V) × 5 (RIO regions: frontal, central, fronto-central, centro-parietal, occipital) revealed a significant main effect of stimulus type, $F(1, 76) = 52.4$, $p < 0.001$, $\eta^2 = 0.734$, indicating that the amplitudes of AV were significant positive than (A+V) stimulus. There was also a significant main effect of RIO regions, $F(4, 76) = 123.5$, $p < 0.001$, $\eta^2 = 0.66$, which showed that occipital region has the strongest difference of amplitudes compared to other four regions. Furthermore, the ANOVA analysis showed a significant interaction between the two factors, $F(4, 76)= 6.352$, $p < 0.001$, $\eta^2 = 0.251$. Further post-hoc comparisons were conducted to reveal the variance of different amplitudes at different regions. The amplitudes of frontal ($p < 0.01$), fronto-central ($p < 0.001$), central ($p < 0.001$), centro-parietal ($p < 0.001$) and occipital site ($p < 0.001$) were more positive in AV (mean potential: -4.1 μV, -1.67 μV, 0.3 μV, 0.34 μV, 0.91 μV) compared with (A+V) (mean potential: -6.01 μV, -4.11 μV, -3.22 μV, -2.35 μV, -0.9 μV).

And at time interval 300-360 ms, the observed integration were different in the frontal (F3, Fz, F4, F8), fronto-central (FC1, FC2, FC5, FC6), the central (C3, C4, Cz), centro-parietal (CP5, CP1, CP2) regions, the ANOVA analysis 2 (stimulus type: AV, A+V) × 4 (RIO regions: frontal, central, fronto-central, centro-parietal) revealed a significant main effect of stimulus type, $F(1, 57) = 46.7$, $p < 0.001$, $\eta^2 = 0.71$, indicating that the amplitudes of AV were significant negative than (A+V) stimulus. There was also a significant main effect of RIO regions, $F(3, 57) = 140.18$, $p < 0.001$, $\eta^2 = 0.88$, which demonstrated that cetro-parietal region has the strongest difference of amplitudes compared to other three regions. Moreover, significant interaction was found between the two factors, $F(3, 57) = 14.78$, $p < 0.001$, $\eta^2 = 0.44$. Further post-hoc comparisons were conducted to reveal the variance of different amplitude at different regions. The amplitudes of frontal ($p < 0.001$), fronto-central ($p < 0.001$), central ($p < 0.001$), centro-parietal ($p < 0.001$) site were more negative in AV (mean potential: -6.65 μV, -4.65 μV, -2.5 μV, 0.02 μV) compared with (A+V) (mean potential: -3.82 μV, -0.58 μV, 1.11 μV, 2.31μV).

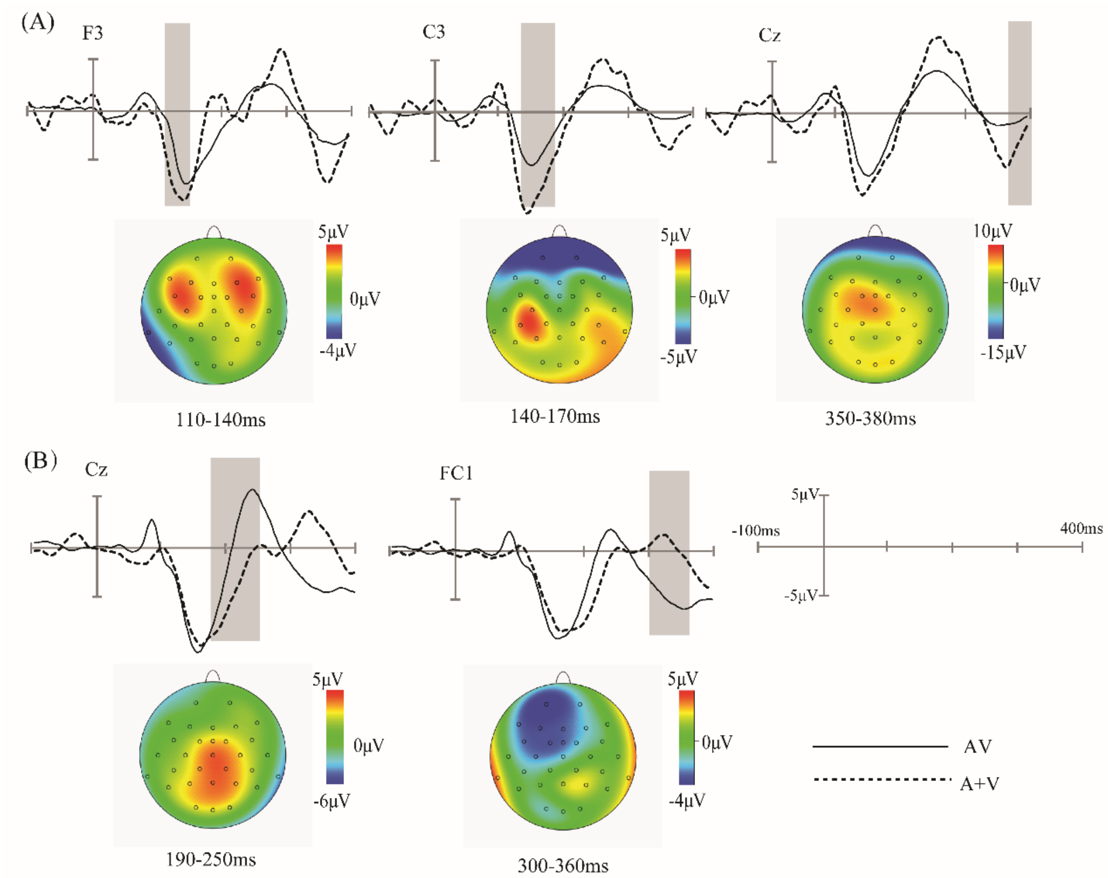


Figure 4. *Grand-average event-related potential and topography map of audiovisual integration elicited by audiovisual stimuli. Event-related potentials of the sum of the unimodal stimuli (A+V) and bimodal stimuli (AV) are shown from 100ms before the stimuli to 400ms after stimulus onset in the different age groups (A) older group, (B) young group. And topography map difference of AV and A+V with significant audiovisual integration in different age groups (A) older group, (B) young group.*

2.4 Discussion

This study investigated audiovisual integration during aging, and the behavioral results revealed that the audiovisual facilitation was greater in the old group when compared to the young group. The results are in line with the research from the simultaneous presentation of semantically congruent sensory stimuli. Laurienti et al. (2006) conveyed that a decline in unisensory performance could be directly attributed to enhanced audiovisual performance, as old adults benefited more than young adults from receiving audiovisual information across multiple sensory modalities. Laurienti et al. (2006) furthered that such enhanced audiovisual gain is likely the factor that old adults could benefit more from combining information from visual and auditory due to the unisensory deficits associated with aging [28]. In our current study, the old group was significantly impaired relative to the young group on the unisensory conditions, suggesting that their unisesnory acuity was lower to young group. Thus, audiovisual gain in the current study is also the result of unimodal modalities decline in old group. Reaction time can offer evidence of the advantage of the audiovisual condition. However, whether this effect implies that the brain effectively utilizes single-modality information, i.e., whether the representation of visual and auditory information under bimodal condition can generate an integration facilitation effect, needs further to be explored. After the redundancy effect occurs, there are primarily two mathematical models for explanation. The race model suggested that during the processing of bimodal stimuli, integration does not occur as a unified process, but rather each stimulus is processed separately within its respective single-modality pathway. The final response of participants is determined by the fastest triggered single sensory modality. To illustrate this point more vividly, taking the example of flipping coins, the probability of getting heads up when flipping two coins simultaneously is much higher than the probability of getting heads up when flipping just one coin. In this way, the occurrence of the redundancy effect is merely a statistical convenience. Conversely, the coactivation model suggests that information from bimodal channels converges at a specific processing stage, and the brain integrates it into a unified perceptual experience. The reliability and stability of this integrated information surpass that of any single-modality information. Therefore, we compared the audiovisual difference between older adults and young adults using race model. Result showed that the violation of race model in young adults was from 370 - 670 ms with a peak 17.98%, while the violation of race model in older adults was from 380 - 740 ms with a peak 26%. Therefore, the results make sure that greater audiovisual integration in older adults was found.

For audiovisual integration at an early stage, it was observed that integration effects in the old group (110-140 ms, 140-170 ms) occurred earlier when compared to the young group (190-250 ms). This is consistent with the study referring to audiovisual speech perception, which investigated age-related differences in the processes underlying audiovisual speech perception [3]. The results showed that the audiovisual interaction was more pronounced at an earlier time point in the old group than the young group. The author contributed such earlier audiovisual interaction to fact that the availability of visual speech cues compensated for less optimal hearing in the old group

[3]. Meanwhile, other compensatory interpretation is often invoked when the old adults show more activity in a brain region than the young group [101], which was proposed as a compensation strategy of the aging process [102,103]. In more details, the old group engaged brain areas to a greater extent than young group to compensate for impaired function in other brain areas [27]. The current study showed that in the old group, relative early and wide integration range (110-140 ms, 140-170 ms) signified those older adults activate an increased brain network in response to cross-modal stimuli, and multisensory integration might play a compensatory role in normal aging.

Additionally, at late audiovisual integration stage, the audiovisual time window was delayed in old group (350-380 ms) when comparing to young group (300-360 ms). Same as in our results, the research of Ren et al. (2018) using an auditory or visual stimuli discrimination experiment showed late audiovisual integration occurred in the 210-240 ms time interval for young adults and the 280-300 ms time intervals in old group [104], suggesting that general reduced cognitive function may be associated with delayed late integration in older adults [9]. Moreover, in the research of Wu et al. (2012) referring to the patients with mild cognitive impairment or Alzheimer's disease, they compared the audiovisual integration of these patients with that of healthy aging participants [105]. The results showed that delayed audiovisual integration and functional deficits related to cerebral atrophy was observed in patients with Alzheimer's disease. While there is a cognitive functional decline with aging [106], therefore, the cognitive functional decline can also be attributed to the delayed audiovisual integration. Besides, with aging, there is a general decline in motor speed and executive function

[107]. Thus, it is reasonable that audiovisual interaction was delayed in older adults.

Chapter 3 The difference in effects of audiovisual integration on working memory between young and older adults

3.1 Background

The combination of information from different sensory modalities plays an important role in generating coherent perception, which is called audiovisual integration. Our brain can take advantage of audiovisual integration to facilitate the individual's performance including the improvement in accuracy and decreased reaction times. Recently, some researchers investigated whether audiovisual integration could induce better training effectiveness to perceptual processing. After comparing effects of audiovisual training and visual training on a coherent motion detection and discrimination task, the audiovisual training group showed greater learning within first session and the subsequent 10 training sessions [108]. Similar results are also found in the research exploring appropriate training on speech identification. Preceding audiovisual training results in better identification when compared with auditory training. These results indicated that audiovisual integration shows benefits on subsequent low-level perceptual processing [109]. The audiovisual benefit on later performance might be related to the fact that audiovisual presentation affects the representation of later perceptual processing. Specifically, some frameworks suggested that audiovisual processing facilitates changes within the later unisensory representations. In other words, audiovisual integration impacts the same structures and representations that are altered during traditional unisensory processing. The result is that later presentation of unisensory stimuli will elicit stronger activation of the unisensory structures. Essentially, audiovisual integration enhances the processing efficiency of the later unisensory information. In contrast to the above framework, the second framework suggests that audiovisual processing involves the alteration of connections between different sensory modalities (visual and auditory) or the formation/alteration of audiovisual representations. In this case, the learning process affects the brain's network of regions involved in processing multiple sensory inputs. As a result, when unisensory stimuli are presented later, they activate a wider network of brain regions that integrate information from multiple modalities. This leads to a more extensive engagement of multisensory processing during unisensory tasks. Notably, the above-mentioned research depicted that audiovisual integration has a positive effect on later perceptual processing. However, the audiovisual integration may also have a positive effect on higher cognitive processing. For instance, participants are better at monitoring and recalling stimuli presented in audiovisual conditions than visual-visual or auditory-auditory stimuli. Moreover, some researchers found higher hit rates for stimuli presented in an audiovisual manner. Two main models have been proposed to explain in which form the information is maintained in working memory: a model of domain-independent store and a multiple-component model. According to the model of domain-independent store, the visual and auditory information from different modalities is maintained in the form of audiovisual integration [86]. This model suggested that auditory stimuli and visual stimuli are integrated in working memory, which could not only provide redundancy of information but might further enable stimuli to be encoded into richer audiovisual representations [86]. In more detail, audiovisual stimuli lead to a stronger trace because additional encoding occurs in multisensory regions. During the later memory processing stage, it can evoke the entire audiovisual representations, resulting in enhanced memory performance. However, Baddeley and Hitch assume that information of working memory is maintained in its respective domain-specific store, followed by the multiple-component model. In more detail, the information is assumed to be stored in two domain-specific subsystems (phonological loop, visuospatial sketch pad) that are governed by the central executive which manipulates the temporary information to complete the current task [110]. Working memory may rely on the processing efficiency of these components. Visual or spatial information was stored in a visuospatial sketchpad, and the acoustic information was stored in a phonological loop. Facilitating the processing of this information should impact the domain-specific stores. The audiovisual stimuli presented in working memory may offer a reliable way to benefit information processing and enhance working memory performance. Generally, we can infer that no matter in which model working memory is engaged, audiovisual benefit can still exist. In addition, there is accumulating evidence that audiovisual stimuli benefit working memory performance [83,111]. When the participants completed an n-back task (0- to 3-back) under visual-only, auditory-only and audiovisual conditions, faster responses and improved accuracy were found during audiovisual condition compared to unisensory conditions [87]. Moreover, in order to explore the neural basis underlying audiovisual working memory, studies compared semantically related audiovisual stimulus presentations with unimodal stimulus presentations using the standardized low-resolution brain electromagnetic tomography (sLORETA) [88]. The results found that the ERP for audiovisual stimuli differed from the ERP for the sum of auditory and visual stimuli at 236 - 530 ms over the frontal and parietal-occipital regions, suggesting that specific brain regions activities in audiovisual working memory [88]. Their subsequent study investigated the oscillation characteristics of neural signals when participants encoded visual, auditory, or audiovisual object in working memory tasks. The theta oscillation was significantly greater in cortical regions including prefrontal, parietal, temporal, and occipital cortices during audiovisual object encoding compared with single modality object encoding [89]. These studies reflect that working memory that incorporates audiovisual information, would not only provide redundancy information but also induce greater activity of brain regions to facilitate working memory performance. However, the above-mentioned working memory task was based on a delay-sample task, which included 3 stages: the encoding stage in which participants need to hold the information in mind, the stage of maintenance (a short delay for participants to keep the information), retrieval stage in which participants need to figure out whether current stimuli matched the stimuli presented in encoding stage. Notably, n-back task needs to monitor and update the consistent information and manipulate the information, which was assumed to be demanding. Here may raise the question that whether audiovisual facilitation on n-back could be observed. In addition, working memory abilities tend to decline with age. The audiovisual integration may have a positive influence on working memory, which helps older adults keep their working memory capacity. It's well established that attention, executive functions and sensory undergo substantial changes during aging, which may influence the audiovisual integration. We have found that audiovisual integration in older adults was greater in young adults, as presented in *Experiment* 1. Here may raise the question that whether audiovisual integration has a positive effect on working memory in older adults.

The first purpose of the present study was to figure out whether audiovisual integration leads to better working memory performance. In order to achieve this, we designed audiovisual n-back, visual n-back, auditory n-back tasks. The basic logic is that the working memory performance is greater in an audiovisual n-back task compared to the working memory performance under visual n-back or auditory n-back tasks. The second purpose is to figure out which age group’s working memory performance is more affected by audiovisual integration.

3.2 Methods

3.2.1 Participants

Twenty-one young adults (10 females, mean age = 21.3 years, SD = 1.9) and 20 older adults (11 females, mean age = 63.5 years, SD = 3.1) were recruited for the present study. All participants were right-handed and reported having normal or corrected-to-normal vision and hearing capabilities. In addition, the Mini-Mental State Examination (MMSE) was used to screen for cognitive impairments in the older participants. This study was approved by the Ethics Committee of Hubei University, and informed consent was acquired from each participant. All participants were given a reward after the procedures were completed.

3.2.2 Stimuli and Procedure

The visual stimuli we used in the task were from Snodgrass and Vanderwart [112], including the animals images with black- and white line. The auditory stimuli were animals' sounds selected from website (http://www.findsounds.com).

Each participant completed the n-back task under three conditions: auditory, visual, audiovisual conditions. The order of auditory, visual, and AV conditions was counterbalanced across participants. In the auditory condition, only animal's sounds were presented from earphones for 500 ms, followed by an interstimulus interval of 2500 ms, while the visual image was not presented. In the visual condition, the animal's images were also presented for 500 ms, followed by an interstimulus intervals of 2500 ms, while the auditory sitmuli were not presented. In the audiovisual condition, the visual and auditory stimuli were presented simultaneously for 500 ms, followed by an interstimulus interval of 2500 ms. The task was based on n-back paradigm, no matter in which condition, the participants were asked to decide whether the currently presented stimuli matched the stimuli presented in the previous trial (1-back condition), 2 trials before (2-back condition), or 3 trials before (3-back condition). If the currently presented stimuli matched the stimuli presented in the previous trial, the participants were required to press "A" button on the keyboard (target). On the other hand, if the currently presented stimuli did not match the stimuli presented in the previous trial, the participants were required to press "L" button on the keyboard (un-target). In each condition, the matched trials accounted for 30% of the total trials, whereas the unmatched trials accounted for 70% of the total trials.

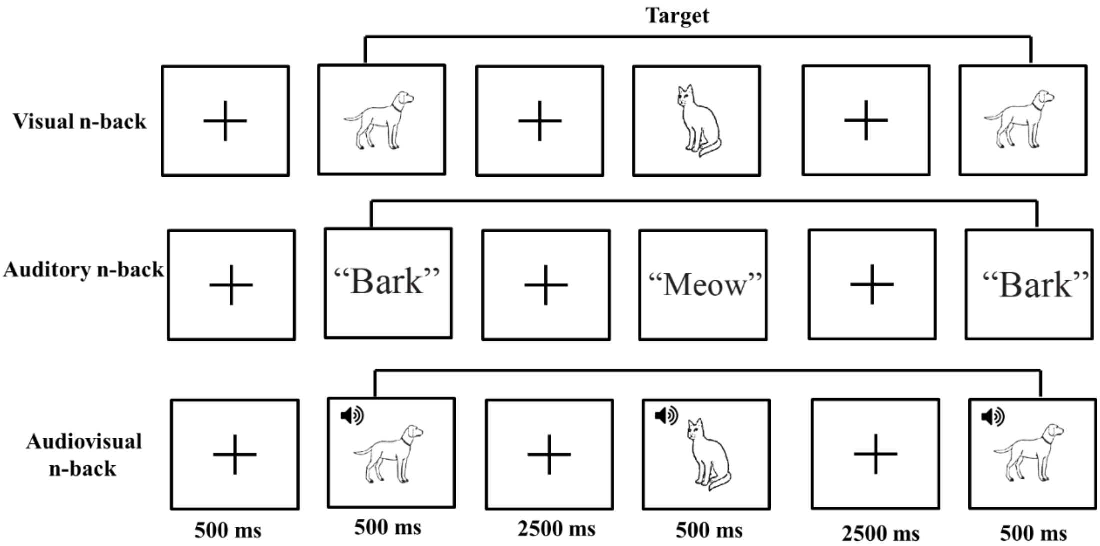


Figure 1. *The visual, auditory, audiovisual n-back tasks were used, illustrated for a 2-back condition.*

3.2.3 Data analysis

In order to explore the behavioral differences in visual, auditory, audiovisual n-back conditions across young and older adults. Repeated-measures ANOVAs with tasks (audiovisual n-back, visual n-back, auditory n-back) × load (1-, 2-, 3-back) × group (older group, young group) were performed separately for reaction time (RT) and accuracy in the training task. Bonferroni-corrected were used in pairwise comparison with post-hoc tests.

In many cognitive tasks, there exists a trade-off between speed and accuracy. This means that individuals need to make a choice between pursuing faster completion speed and higher accuracy when performing tasks. Generally, improving speed might lead to a decrease in accuracy, while aiming for higher accuracy might extend the time taken to complete the task. This trade-off between speed and accuracy is quite common in various cognitive tasks. For instance, in visual search tasks, individuals might face the choice between quickly locating a target with a potential for errors, or taking more time to accurately find the target. In working memory tasks, individuals might balance maintaining information accuracy with processing information more swiftly. The Inverse Efficiency Score (IES) is an index used to measure performance in cognitive tasks. It combines both task accuracy and the time taken to complete the task to provide a more comprehensive evaluation of an individual's performance in the task. Typically, the IES is calculated by dividing the task completion time by task accuracy, yielding an integrated index to assess the time taken for a given level of accuracy. A lower Inverse Efficiency Score indicates that an individual exhibits both high accuracy and efficiency in task performance. In other words, individuals with lower Inverse Efficiency Scores complete tasks in relatively shorter time periods while maintaining the same level of accuracy, indicating greater performance. This index assists researchers in comparing different individuals or tasks, enabling a more comprehensive understanding of their cognitive performance. Repeated-measures ANOVAs with tasks (audiovisual n-back, visual n-back, auditory n-back) × load (1-, 2-, 3-back) × group (older group, young group) were performed for IES.

In multisensory integration quantification, researchers also investigate the contributions of visual or auditory modalities to the multisensory integration effect. The contribution of visual modality can be represented as VE = (AV - A) / A, where VE stands for the contribution of visual modality to audiovisual integration. AV represents the accuracy under audiovisual conditions, and A represents the accuracy under single auditory conditions. Similarly, the contribution of auditory can be expressed as AE = (AV - V) / V, where AE represents the contribution of auditory modality to audiovisual integration. V represents the accuracy under single visual conditions.

3.3 Results

3.3.1 Hit rates

The analysis showed that there was a main effect of tasks, $F(2, 76) = 29.056$, $p < 0.001$, $\eta^2 = 0.433$, indicating that the accuracy at audiovisual n-back was higher than that of visual n-back task and auditory n-back task. The main effect of load was also significant, $F(2, 78) = 68.49$, $p < 0.001$, $\eta^2 = 0.674$, indicating that the accuracy of 1-back was significantly greater than that of 2-back, 3-back. The main effect of group was significant, $F(1, 38) = 101.359$, $p < 0.001$, $\eta^2 = 0.727$, indicating that the accuracy in young adults was greater than that of older adults. In addition, the interaction between task and group was significant, $F(2, 76) = 11.508$, $p < 0.001$, $\eta^2 = 0.232$. For older adults, post hoc analysis showed the accuracy of the audiovisual n-back task is significantly higher than that of the visual n-back and auditory n-back tasks. However, among young adults, there is no significant difference in accuracy across the three types of tasks. From a task perspective, under the visual task condition, the accuracy of young adults remains higher than that of older individuals, regardless of n-back levels. The accuracy in the auditory condition and the audiovisual condition shows a similar difference compared to the accuracy in the visual task condition. We also found an interaction between load and group, $F(2, 76) = 34.776$, $p < 0.001$, $\eta^2 = 0.478$. The post hoc analysis showed that the accuracy gradually decreases as the load increases in older adults. However, such a phenomenon was not found in young adults. The interaction between load and task was found, $F(4, 152) = 1.569$, $p = 0.015$, $\eta^2 = 0.078$. At the 1-back level, the accuracy in audiovisual n-back was significantly greater than that of visual and auditory n-back. At the 2-back and 3-back levels, the audiovisual benefit was still found. No other significant main effects or interactions were found, *ps* > 0.05.

3.3.2 Response times

The analysis showed that the main effect of tasks was significant, $F(2, 76) = 24.583$, $p < 0.001$, $\eta^2 = 0.393$, suggesting that the reaction time of audiovisual n-back task was faster than that of visual and auditory n-back tasks. Moreover, the main effect of the load was also significant, $F(2, 76) = 100.84$, $p < 0.001$, $\eta^2 = 0.726$, suggesting that the reaction time of 1-back was faster than that of 2-back and 3-back. The interaction between task and group was significant, $F(2, 76) = 3.238$, $p = 0.045$, $\eta^2 = 0.079$. The post-hoc analysis indicated that the reaction time of audiovisual task were significantly shorter than that of auditory task and visual task in older adults. In addition, in young adults, the reaction time of audiovisual tasks was significantly shorter than that of auditory tasks, similar as older adults. In terms of tasks, significant differences were found between older adults and young adults at visual n-back conditions. That is, the reaction time of young adults was significantly shorter than that of older adults. Moreover, the interaction between task, load, and group was significant, $F(4, 152) = 2.789$, $p = 0.028$, $\eta^2 = 0.068$. The post-hoc analysis found that, for older adults, the reaction time was significantly shorter in audiovisual n-back compared to visual n-back at the 1-back level. The faster reaction time was also found in audiovisual n-back compared to auditory n-back at the 2-back level. Notably, faster reaction time were observed in audiovisual n-back compared to visual and auditory n-back at a 3-back level. For young adults, the reaction time was significantly faster in audiovisual and visual n-back tasks compared to auditory n-back tasks at the 1-back level. At the 2-back level, the reaction time of the audiovisual n-back task was shorter than that of the auditory n-back task. At the 3-back level, a similar effect was found. That is, a shorter reaction time was found in audiovisual n-back compared to auditory n-back. In terms of task, we found that the reaction time of young adults was overall shorter than that of older adults across 1-, 2-, and 3-back levels in the visual n-back task. For the auditory n-back task, the reaction time of the two groups were similar at 1- and 3-back levels. The shorter reaction time in young adults was found at the 2-back level. For the audiovisual n-back task, the reaction time in young adults was shorter than that of older adults at the 2-back level. At 1-back and 3-back levels, the reaction time of the two groups were similar.

3.3.3 IES

The analysis for IES has shown that the main effect of tasks was significant, *F*(2,

76) = 41.709, $p < 0.001$, $\eta^2 = 0.523$, indicating that the IES was lower in audiovisual n-back task than these of in visual and auditory n-back. The main effect of the load was also found, $F(2, 76) = 118.368$, $p < 0.001$, $\eta^2 = 0.757$, suggesting that the IES was lower at the 1-back level than these at the 2- and 3-back levels. Moreover, the main effect of group was also observed, $F(1, 38) = 23.864$, $p < 0.001$, $\eta^2 = 0.386$, showing that young adults obtained a lower IES than older adults. In terms of two-way interaction, the interaction between tasks and group, $F(2, 76) = 9.022$, $p < 0.001$, $\eta^2 = 0.192$. For older adults, the IES in audiovisual n-back task was lower than in visual and auditory n-backs. For young adults, the IES in audiovisual and visual n-back were lower than in auditory n-back. The interaction between load and group, $F(2, 76) = 14.045$, $p < 0.001$, $\eta^2 = 0.270$, was significant. Post-hoc analysis showed that the IES decreased as the load increased in older adults. For young adults, similar effect was found with IES decreasing as the load increased. Moreover, the interaction between task and load was found, $F(4, 152) = 14.045$, $p < 0.001$, $\eta^2 = 0.174$. The IES at the 1-back level was lower than at the 2-back and 3-back levels across audiovisual, visual, auditory n-back. In terms of a three-way, the interaction between task, load, group was found, $F(4, 152) = 4.751$, $p = 0.001$, $\eta^2 = 0.111$. For older adults, the audiovisual effect was observed across 1-, 2- and 3-back levels, indicating that the IES was lower in the audiovisual n-back task than in the auditory, visual n-back task. For young adults, the IES was lower in the audiovisual n-back task than in the visual n-back task at 1-back level. At 2-back and 3-back levels, the IES was lower in the audiovisual n-back task than in the auditory n-back task.

3.3.4 Visual and auditory contribution

The analysis showed that the main effect of modality was found, $F(1, 38) = 8.724$, $p = 0.005$, $\eta^2 = 0.187$, suggesting that the contribution from visual modality was greater than the contribution from auditory modality. The main effect of load was found, $F(2, 76) = 4.478$, $p = 0.015$, $\eta^2 = 0.105$, suggesting that the contribution in 3-back level was greater than that in the 2-back level. Moreover, the main effect of group was found, indicating that the overall contribution in older adults was greater in young adults. In terms of two-way interaction, the interaction between load and group was found, $F(2, 76) = 7.549$, $p = 0.001$, $\eta^2 = 0.166$. For older adults, the contribution for 3-back level was greater than 2- and 1-back levels. For young adults, there was no significant difference between 1-, 2 and 3-back levels. The three-way interaction was found. After post-hoc analysis, the results showed that the contribution from visual modality in the 3-back level was greater in 1- and 2-back levels in older adults. However, only a greater contribution from auditory modality was found in 3-back level when comparing to 2-back level in older adults. For young adults, there was no significant difference in the contribution from visual or auditory modality across 1-, 2-, 3-back levels.

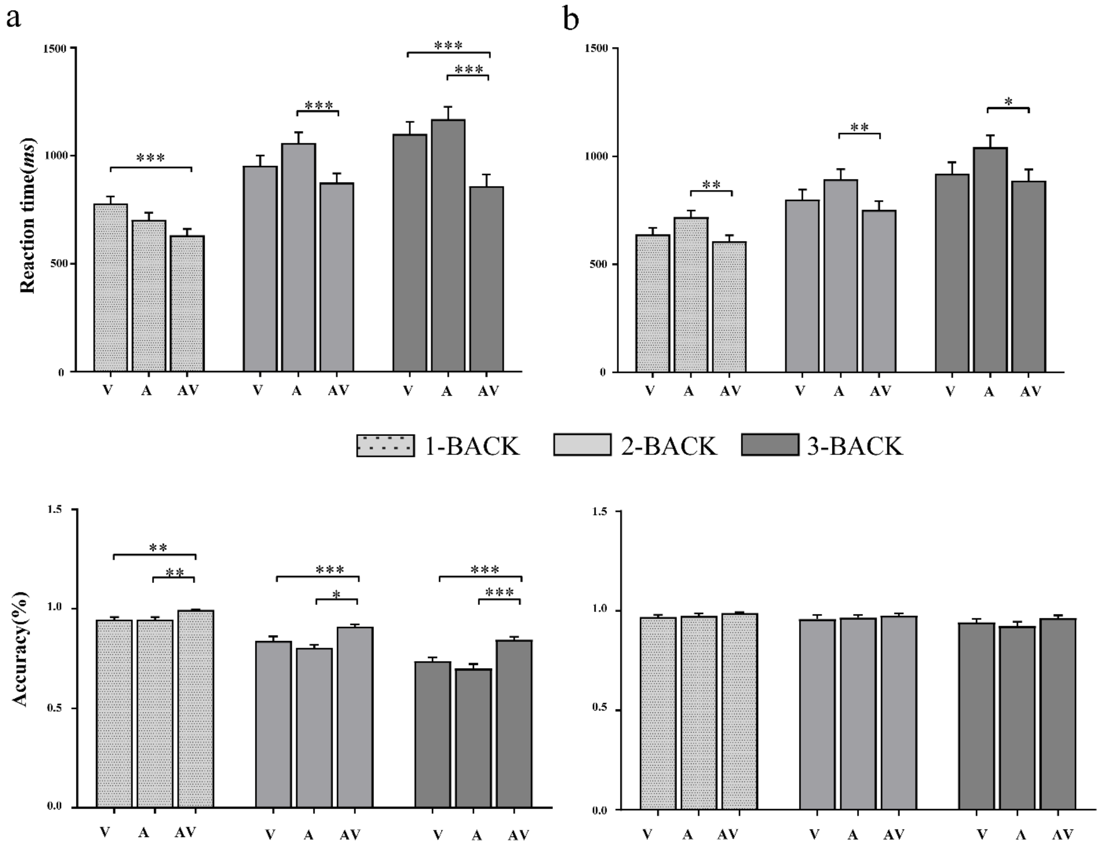


Figure 2. *Differences in response time and accuracy between older adults and young adults. (a) represents the response time and accuracy in older adults while (b) represents these data in young adults.*

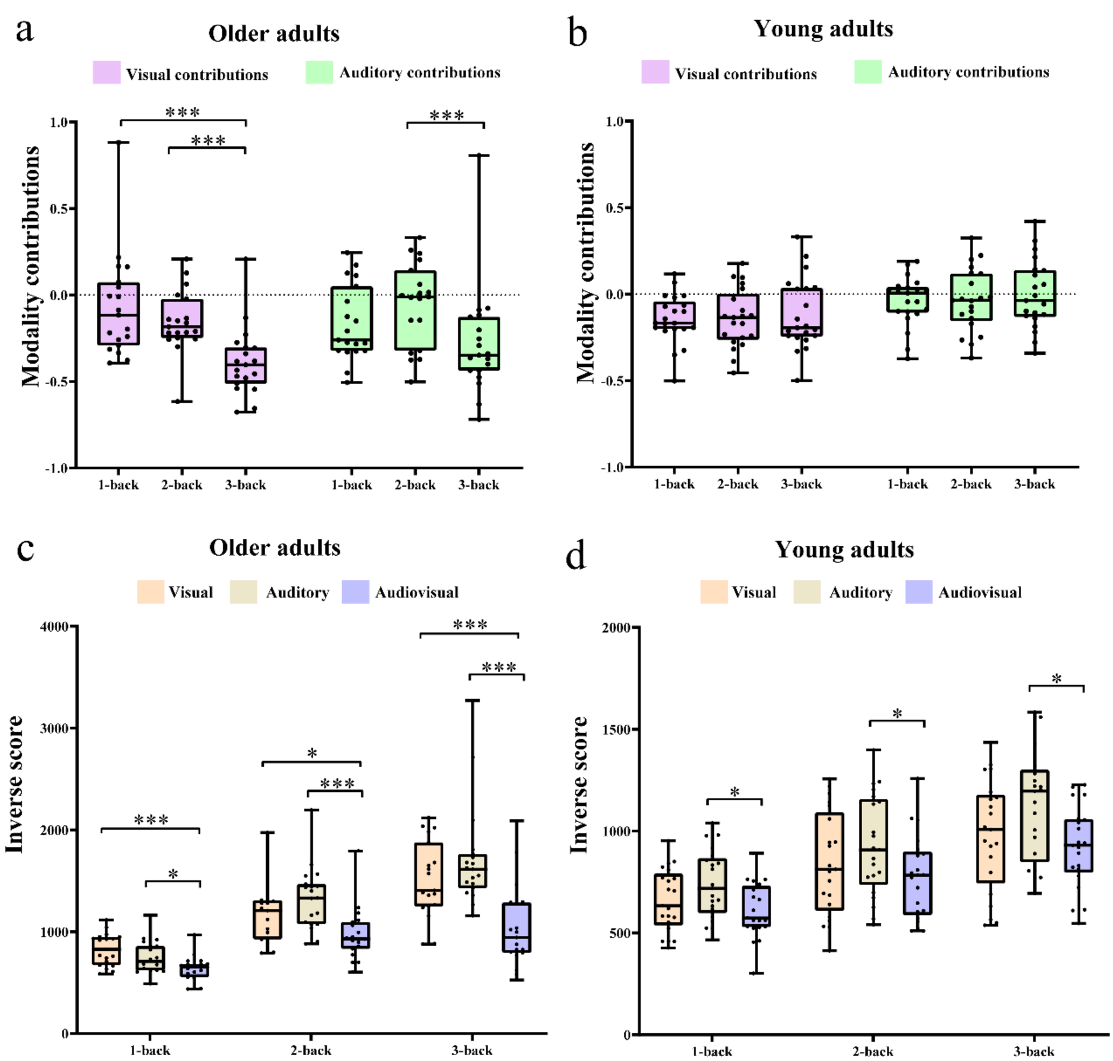


Figure 3. *Differences in modality contributions and IES between older adults and young adults. (a) represents visual and auditory contributions in older adults while (b) represents these data in young adults. (c) represents IES in older adults while (d) represents IES in young adults.*

3.4 Discussion

As expected, the study found that as the working memory load increased, irrespective of the modality being tested, there was a decline in accuracy and an increase in response time (RT). This means that participants experienced greater difficulty in accurately completing the tasks and took longer to respond as the demands on their working memory increased. The observed pattern of results supports the well-established notion that working memory capacity is limited and that surpassing its capacity leads to decreased performance. This pattern of declining accuracy and increasing response time with higher working memory load was consistent across both young and older adults, indicating that age did not significantly influence the observed effects. These results suggest that the working memory limitations observed are applicable to individuals across different age groups. In addition, the effectiveness of the n-back manipulation in taxing working memory resources. The n-back task, commonly used in cognitive psychology research, requires participants to identify whether the current stimulus matches the one presented "n" steps back in a sequence. By systematically increasing the "n" level, the researchers were able to progressively load participants' working memory capacity. The observed decline in accuracy and increase in response time with each successive increase in working memory load, regardless of the modality being tested, further supports the validity and reliability of the n-back manipulation as a tool for assessing working memory performance. In addition, in consistent with previous literature, compared with young adults, older adults performed more slowly across 3-back working memory loads and less accurately under certain WM loads (1-back, 2-back, 3-back).

Our research aimed to investigate the influence of audiovisual processing on subsequent working memory performance in terms of audiovisual facilitation. The obtained results confirmed our predictions. That is, the presentation of audiovisual stimuli in working memory led to improved working memory performance compared to the presentation of visual and auditory alone. According to the conceptual short-term memory (CSTM), short-term memory is not simply about holding information in a limited storage, but rather involves integrating new perceptual information with existing long-term memory. Specially, when individuals perceive meaningful stimuli, these stimuli are rapidly identified at a conceptual level, activating relevant long-term memory information. Subsequently, activated concepts form new connections with each other, giving rise to novel representations that are encoded into long-term memory [113]. These new representations can be retained within working memory or consolidated into long-term memory. The core idea of the CSTM model is that individuals not only maintain information in short-term memory but also integrate it with existing long-term memory, thereby generating more elaborate and meaningful representations. This integration can expedite information processing and storage, facilitating easier retention and retrieval of information in memory. Therefore, during working memory processing, the presentation of audiovisual stimuli leads to rapid recognition at the conceptual level and activates associated long-term memory information. This, in turn, facilitates the maintenance and retrieval of information within working memory. Moreover, the memory trace theory suggests that when information from different sensory modality is integrated, multisensory regions become activated. Activation of these multisensory regions forms more stable memory traces, with subsequent cognitive processing continuing to activate these memory traces. This enables these cognitive processes to operate on richer and more informative representations, thereby enhancing working memory efficiency [114]. In this study, an audiovisual n-back task was employed. From a cognitive neuroprocessing perspective, this task encompasses encoding, maintenance, and updating processes [115]. The encoding stage is crucial for working memory operations and significantly influences subsequent memory processes such as maintenance and updating. During audiovisual working memory processing, the audiovisual memory traces is activated and become more stable when individuals encode audiovisual stimuli. Simultaneously, subsequent processes like maintenance and updating are facilitated due to the activation of these visual-auditory memory traces.

Moreover, in terms of IES, the results revealed that in older adults, the IES under the audiovisual n-back condition was smaller compared to the visual n-back and auditory n-back conditions across 1-, 2- and 3-back levels. However, such an effect was not found in young adults. These results may indicate that older adults were more adept at utilizing audiovisual integration compared to young adults. Previous studies have shown that the audiovisual integration in older adults is greater than in young adults. This may be because older adults need to take advantage of audiovisual integration to compensate the decline in sensory and cognitive abilities. Some previous studies have suggested that audiovisual integration with meaningless stimulus occurs at a lower-level perceptual stage, whereas audiovisual integration with meaningful stimulus occurs at a higher-level cognitive stage. However, lower-level perceptual processing does not necessarily lead to increased audiovisual integration, while higher-level cognitive processing triggers compensatory mechanisms for enhanced audiovisual integration. Specifically, during the processing of meaningful stimuli, older adults exhibit enhanced audiovisual integration to compensate for their attentional decline

[116].

Notably, greater visual contribution and auditory contribution were found in the 3-back level when comparing to 1-back, 2-back levels in older adults. The available explanation was related to inverse effectiveness. The inverse effectiveness suggested that the audiovisual integration is the most effective when processing the ambiguous stimuli which may pose a certain level of difficulty. Some Researchers explore multisensory integration when processing degraded and challenging multisensory information. The study explores the advantages of multisensory integration in object recognition under noisy or degraded stimuli and investigates the associated neural mechanisms. The results proved the inverse principle, which suggests that the effects of multisensory integration are more pronounced when stimuli are degraded or ambiguous. Two experiments were conducted to compare the processing of degraded and clear multisensory information at both behavioral and neural levels, with a particular focus on the roles of the temporal gyrus, parietal operculum, and intraparietal sulcus (IPS) brain regions. The research employed the Bayesian hierarchical drift-diffusion model (HDDM) and dynamic, degraded stimuli to address these questions, while also emphasizing the role of IPS in inverse effectiveness. That is, when dealing with higher difficulty tasks, the multisensory network in the fronto-parietal cortex, particularly the IPS region, becomes activated during multisensory integration. In contrast, during the integration of multisensory stimuli in lower difficulty tasks, direct information exchange between the visual and auditory cortices can take place. In other words, the inverse effectiveness is attributed to the additional activation of the IPS area during the processing of high difficulty tasks, which leads to multisensory integration information exchange [117]. In our research, the 3-back level requires more cognitive resources than 1- and 2-back level, indicating that inverse effectiveness in this level may occur. The inverse effectiveness may be manipulated by IPS region.

Finally, these results, consistent with previous research, suggest that despite age-related sensory decline, older adults can still derive benefits from audiovisual information. The combination of auditory and visual cues in the AV condition likely compensated for age-related sensory impairments and enhanced the older adults' ability to process information efficiently. What sets this study apart is the unique finding that the early multisensory benefit from AV presentations improved the subsequent working memory performance of older adults. The audiovisual stimuli seemed to facilitate both the speed and accuracy of the older adults' working memory performance to the extent. This suggests that the audiovisual integration provided a perceptual advantage that not only enhanced working memory performance but also mitigated age-related differences in working memory abilities. The combination of auditory and visual information likely improved encoding, retention, and retrieval processes, leading to more efficient working memory performance in older adults. Overall, these findings highlight the potential of audiovisual stimuli to enhance cognitive performance in older adults. Understanding the benefits of audiovisual integration can inform the development of interventions and technologies aimed at improving cognitive functioning, particularly in tasks involving working memory, for individuals of different age groups.

Chapter 4 Comparison for young and older adults: audiovisual n-back training effect on working memory performance

4.1 Background

Working memory is a cognitive system used for the temporary maintenance and manipulation of information [41]. Modality-specific working memory training induces changes in neural activation and improves working memory performance. Some reports have shown that visual n-back training leads to enhanced prefrontal and parietal activations responsible for visual working memory storage [54]. A similar activation difference was found in auditory working memory training, in which the right inferior frontal regions related to maintenance of auditory information were engaged in the improvement of working memory [57]. The differences between modality-specific working memory training were further explored using functional magnetic resonance imaging (fMRI). Following two weeks of separate visual and auditory n-back training, researchers examined the transfer effects of these two types of working memory training on an untrained visual working memory task. Only visual n-back training induced additional activation in the right middle frontal regions on the untrained visual task; such decreased activation was not observed after auditory n-back training. The authors further confirmed that the right middle frontal regions are specific to the maintenance and manipulation of visual information [58]. These studies have suggested that modality-specific training produces specific activation changes in the working memory network. However, according to the multiple-component model, working memory consists of a central executive and two components specialized for maintaining modality-specific information [41]. The phonological loop is specialized for retaining auditory and phonological information, while the visuospatial sketchpad is used to retain visual spatial and nonspatial information. Working memory relies on the processing efficiency of the visuospatial sketchpad and the phonological loop [87]; thus, facilitating the storage and processing of these components may enhance working memory performance [57,58].

Dual n-back training is in line with a multiple-component model, requiring the simultaneous processing of visual and auditory stimuli. The visual stimuli consist of blue squares in eight different locations and presented one by one. The auditory stimuli consist of the sound of single letters. Working memory contents are monitored and updated in these two modalities separately. There has been considerable interest in the dual n-back task given its potential to improve working memory and the transfer effects on fluid intelligence, executive function and attention [71]. Because this form of training engages processes needed to handle incongruent information from visual and auditory modalities, it is not conducive to the involvement of automated processing and task-specific strategies [71]. However, conflicting results regarding dual n-back training have been reported. That is, some researchers have found that this task may not promote a training effect because employing incongruent visual and auditory information leads to competition between modalities and interfere with the participant’s response [93]. For example, in some dual n-back training tasks, greater training and transfer effects were not observed [60,118].

Notably, visual and auditory information may be related, leading to an audiovisual facilitation effect [119]. This effect demonstrates the merging of the related visual and auditory stimuli is integrated into a coherent percept and leads to enhanced information processing at both the behavioral and neural levels [7,14]. That is, when participants memorized audiovisual, visual and auditory stimuli respectively, the accuracy and responses time of working memory performance were overall better during audiovisual stimuli presentation relative to unimodal stimuli presentation in working memory [87]. Subsequent ERP research was conducted to investigate the neural changes associated with audiovisual processing in working memory. The results found that the latency of the P3 component evoked during audiovisual stimuli presentation was earlier than the latency evoked during unimodal stimuli presentation in working memory. The earlier latency indicated faster cognitive processing during audiovisual stimuli presentation in working memory [87]. Cognitive load theory argues that the capacity to handle information is increased by using both visuospatial sketchpad and phonological loop components [120]. When all information has to be processed by the visuospatial sketchpad, its capacity is easily overloaded. The simultaneous presentation of related auditory information may reduce some of the load on the visuospatial sketchpad by shifting it to the phonological loop, thereby enhancing working memory performance [120]. Considering the audiovisual advantage in working memory processes, some researchers have examined the behavioral impact of working memory training with audiovisual stimuli on working memory performance [121]. Two groups of participants completed audiovisual and visual n-back training. The audiovisual n-back training consisted of related visual and auditory object stimuli, such as the image of a bell paired with the sound of a bell. The unimodal n-back training consisted of different visual object stimuli. Their results indicated that the group with audiovisual n-back training not only exhibited equal training gain but also potentially exhibited transfer effects on a complex working memory span task compared to the unimodal n-back training group. However, this research did not address the neural effect of working memory training with audiovisual stimuli [121].

With regard to neural effects, training effects on certain brain regions have been found in previous studies, including activation changes in frontal and partial regions and connectivity between the prefrontal-parietal network [54,122]. Although it is critical to identify specific regions that are influenced by working memory training, fMRI may not fully reveal subcomponents of working memory processing affected by training. Recent studies with ERPs have demonstrated that visual and dual n-back working memory training modulate separate components of the working memory process [76]. They found greater N2 amplitudes, which is related to mismatch/match identification [123]. The subsequent P3 amplitude, which indexes working memory updating, was also enhanced after training [70]. This raises the question of whether working memory training with audiovisual n-back could influence these ERP components (N200, P300) and enhance working memory performance. Previous studies have explored associated brain activity between different processing stages in a delayed matching working memory task [124]. They found that similar activation between early encoding and later maintenance in the lateral prefrontal cortex and visual cortex was related to improved working memory performance. The results indicate that similar brain activity in different processing stages of working memory underlies improved working memory [124]. In this study, early audiovisual processing at the encoding stage and later mismatch identification and updating were involved in audiovisual n-back training task. Audiovisual processing has similar brain activation in the frontal region as mismatch identification and similar brain activation in the central region as updating respectively [123,125,126]. This suggests that working memory with audiovisual stimuli may induce successful working memory performance. According to the information-degradation hypothesis, sensory processing and higher-order cognitive processing share limited cognitive resources [127]. Under this hypothesis, changes in one of the systems can influence the efficiency of the other; that is, dedicating too many cognitive resources to perceptual processing may leave insufficient resources for later higher cognitive processing, such as mismatch identification and working memory updating. The presentation of audiovisual stimuli at the perceptual stage may release more resources for later working memory processing and enhance processing efficiency. Therefore, we hypothesize that audiovisual working memory training may induce enhanced N2 and P3 components, thereby enhancing working memory performance.

Despite the evidence that working memory training improves a variety of cognitive functions, the training-induced transfer effect is still debated. Recent studies have suggested that transfer to untrained working memory tasks may be consistently observed, but the transfer effect on fluid intelligence is not preserved [72,121,128,129]. This conflict may result from ambiguity over whether there is an overlap in the brain regions involved in training and transfer tasks. In particular, some neuroimaging studies have found that working memory updating training yields transfer to an n-back task, in which the updating process is engaged, but no transfer to a Stroop task, which involves the inhibition process [77]. Further investigation of neural activation showed that the overlap of striatal activation between training and the n-back task determined the transfer, indicating that a transfer effect is expected if the training and transfer tasks have specific overlapping brain regions [77]. Some neural evidence has further demonstrated that parts of working memory circuity, such as the left intraparietal region, have been linked to audiovisual processing [130], indicating that the transfer effect of working memory training on audiovisual processing is potentially possible. Moreover, the stronger the correlation between working memory and this other cognitive ability is, the larger the transfer effect expected [131]. Audiovisual processing research has shown that there is a connection between sensory perception and cognitive abilities, reflected by participants with larger working memory capacity exhibiting better performance in audiovisual processing [132]. Therefore, we hypothesized that training that successfully improves working memory would also directly affect audiovisual processing.

To investigate the training effect and transfer effect, we designed an audiovisual n-back working memory task that included related visual and auditory information. Before (pretest) and after training (posttest), the P300 and N200 components were elicited by an audiovisual n-back task with 3 levels (1-, 2-, or 3-back) in both young and older groups. By comparing the P300 and N200 components between pretest and posttest, whether working memory performance could be improved was determined. By testing transfer effects, the current study verified whether working memory training facilitated audiovisual processing using an audiovisual discrimination task.

4.2 Methods

4.2.1 Participants

Thirty-seven healthy young adults and 38 healthy older adults from Hubei University were recruited in current study. For younger adults, they were randomly assigned to a passive control group (17 subjects: mean age = 20.65 years, SD = 1.7) and an audiovisual n-back training group (20 subjects: mean age = 20.52 years, SD = 1.9). For older adults, they were randomly assigned to a passive control group (18 subjects: mean age = 63.3 years, SD = 1.8) and an audiovisual n-back training group (20 subjects: mean age = 65.5 years, SD = 2.3).

All the participants reported having normal or corrected-to-normal vision and hearing abilities. The two groups (young training group vs. young control group, older training group vs. older control group) were comparable in terms of education and fluid intelligence (Raven's Advanced Progressive Matrices).

4.2.2 Procedure

The training group took part in 10 training sessions containing audiovisual n-back tasks over 2 weeks (5 training sessions per week), whereas the control group underwent no training during this time. Each training session was approximately 50 mins, giving about 8 hours of total training dose. At pre- and posttest, participants were required to perform the audiovisual n-back task (1-, 2-, or 3-back) and discrimination task, and ERPs data was collected. The general procedure can be found in **Figure 1b**.

4.2.2.1 Training task

The adaptive audiovisual n-back task was designed to promote learning (see **Figure 1a)**. This task consisted of simultaneously presented visual and auditory stimuli. The visual stimuli were black-and-white line images of animals chosen from Snodgrass and Vanderwart [112], and the auditory stimuli were the corresponding animal sounds selected from a website (http://www.findsounds.com). The visual and auditory stimuli were presented simultaneously for 500 ms, followed by an interstimulus interval of 2500 ms. Each training session began with the 1-back condition. The task was based on the adaptive principle: if participants provided correct responses on at least 90% of trials, the task advanced to the next level (e.g., from 1-back to 2-back). On the other hand, if participants provided correct responses to ≤ 80% of trials, the task difficulty was reduced (e.g., from 2-back to 1-back). Participants responded by pressing the left mouse button when the current stimulus matched the one presented n steps back, or the right mouse button when the current stimulus did not match the one presented n steps back. Each training session contained 20 blocks. Each block consisted of 20 + n trials with 6 targets (matched trials) and 14 nontargets (unmatched trials). After training in each block, participants received feedback on their performance.

4.2.2.2 Training and transfer outcome

At pre- and posttest, participants were instructed to perform a modified version of the audiovisual n-back task. This task provided no feedback, presented the same stimuli as in the training task, was not adaptive (only three n-back conditions: 1-back, 2-back, and 3-back) and was adopted to measure the training effectiveness using EEG and behavioral data. This task consisted of 15 blocks (5 blocks per level).

The subjects also performed an audiovisual discrimination task including target and standard stimuli at both pre- and posttest. The source of the stimuli was the same as that in the training task. In the experiment, the stimulus types consisted of target and standard stimuli. The target stimuli included visual target (the image of dog), auditory target (the sound of dog), and audiovisual target (the presentation of the visual and auditory target simultaneously). The standard stimuli included visual standard stimuli (the image of cat), auditory standard stimuli (the sound of cat), and the audiovisual standard stimuli (the presentation of the visual and auditory target simultaneously). During the experiment, 3 blocks were conducted. Each block contained 36 target stimuli (12 auditory, 12 visual and 12 audiovisual) and 144 standard stimuli (48 auditory, 48 visual and 48 audiovisual). The stimuli were randomly presented for 500 ms with an interstimulus interval of 1,000 ms. Participants were instructed to fix their gaze on the center of the computer screen in which the stimuli were presented. They were asked to press the left mouse button as quickly and accurately as possible when target stimuli were presented.

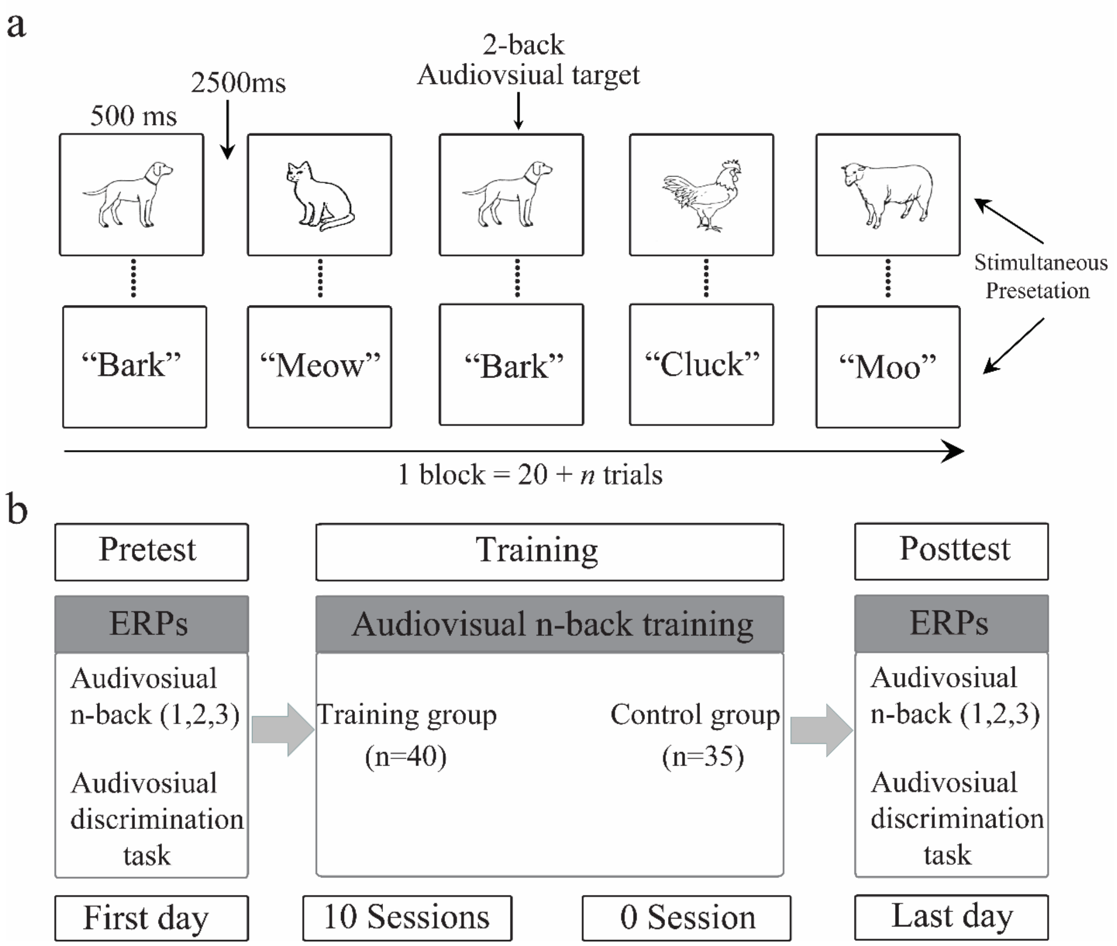


Figure 1. *(a) Example of a 2-back task in the audiovisual working memory training. Each stimulus was presented for 500 ms and contained both auditory and visual information (i.e., participants could both hear and see animals). Participants were instructed to determine whether the currently presented animal matched the animal presented in the previous 2 trials. (b) Schematic description of the study design. All groups performed the audiovisual n-back task (1-, 2-, and 3-back) and audiovisual discrimination tasks at pretest and posttest (ERPs). During training, the training group participated in an adaptive audiovisual n-back task, whereas the control group did not receive any training.*

4.2.3 EEG Data recording and Preprocessing

4.2.3.1 Audiovisual n-back

The EEG activity was recorded with an EEG system (BrainAmp MR plus, Gilching, Germany) using a 32-electrode EEG cap (Easy-cap, Herrsching Breitbrunn, Germany). The reference electrode was positioned at FCz. The horizontal electrooculogram (EOG) data was monitored with an electrode positioned at the outer canthi of the left eye, and the vertical EOG was monitored with an electrode positioned roughly 1 centimeter below the right eye. All the signals were digitized with a sampling rate of 1000 Hz. During EEG recording, the impedances of all electrodes were kept below 5 kΩ.

The offline analysis for EEG data was conducted using functions in EEGLAB and ERPLAB under MATLAB software (2016a). The position of EEG electrodes was based on the 32-channel montage of the international 10-20 system. All data were rereferenced to the mastoid electrodes (TP9, TP10) and were bandpass filtered from 0.1 to 30 Hz at a downsampling rate of 500 Hz. For younger adults, the ERPs elicited by matched trials were divided into epochs starting from 200 ms pre-stimulus to 1000 ms post-stimulus (600 points), with baseline correction made from 200 ms pre-stimulus. For older adults, the ERPs elicited by matched trials were divided into epochs starting from 100 ms pre-stimulus to 600 ms post-stimulus (350 points), with baseline correction made from 100 ms pre-stimulus. The subsequent averaging rejected epochs with large artifacts if the voltage exceeded ± 100 μV. Then, all the remaining trials were averaged separately for different n-back condition (1-, 2-, and 3-back), and ground-averaged data were also acquired by averaging each electrode under each n-back condition across all participants. For N200 and P300 components, peak detection was performed on averaged waveforms of each participant. Based on previous studies, the amplitude of N200 component was regarded as the peak that occurs at around 200 - 350 ms, whereas the amplitude of P300 component was quantified as the peak that occurs at approximately 300 - 600 ms [74,133]. These ERP components amplitude was averaged across electrodes within each brain region, which may reduce the noise resulting from unstable of individual electrodes [74].

4.2.3.2 Audiovisual discrimination task

The EEG recording procedure for the transfer task was identical for the audiovisual n-back task.

The offline analysis for EEG data was conducted using Brain Vision Analyzer software. All data were rereferenced to the mastoid electrodes (TP9, TP10) and were bandpass filtered from 0.01 to 60 Hz at a downsampling rate of 500 Hz. The ERPs elicited by standard trials were divided into epochs starting from 100ms pre-stimulus and 800 post-stimulus (450 points), with baseline correction made from 100 ms pre-stimulus. When the voltage exceeded ± 100 μV, Epochs were considered as to be contaminated with large artifacts. Then these epochs were rejected. The remaining epochs were averaged for each stimulus condition (auditory, visual, and audiovisual), and ground-averaged data were also acquired by averaging each electrode under each stimulus condition across all participants.

4.2.4 Data Analysis

We analyzed the effect of training by examining behavioral differences (accuracy and reaction time) between pretest and posttest. Repeated-measures ANOVAs with session (pretest, posttest) × load (1-, 2-, 3-back) × group (training group, control group) were performed separately for reaction time (RT) and accuracy in the training task. To assess accuracy, the sensitivity index (*d'*) was calculated for every participant and averaged over stage of pretest and posttest. *d'* was estimated as the difference between the hit rate and false-alarm rate. The reaction time was calculated as the average of the sum of reaction times on correct responses to match trials and correct responses to unmatched trials. To compare the training gain among different working memory loads, we also conducted repeated-measures ANOVAs with load (1-, 2-, 3-back) as the within-subject factor and group (training group, control group) as the between-subject factor. The dependent variables were difference in reaction time and accuracy between the posttest and the pretest. For the transfer effect, repeated-measures ANOVAs with session (pretest, posttest) × modality (V, A, AV) × group (training group, control group) were conducted in accuracy and reaction time. Accuracy was the proportion of correct responses to target stimuli relative to total target stimuli. Reaction time was based on the correct responses to target stimuli.

To examine neural changes after training, we selected the following regions of interest (ROIs) based on previous studies [134]: the frontal (F3, Fz, F4), central (C3, Cz, C4), and parietal (P3, Pz, P4) regions. The amplitude of the ERPs components (N2, P3) was averaged across electrodes within each ROI, which may reduce the noise resulting from unstable of individual electrodes [74]. For each ERPs component, repeated-measures ANOVAs were conducted separately to examine amplitudes in each region. The content of the analysis included (pretest, posttest) × group (training, control) × load (1-back, 2-back, 3-back).

To obtain transfer outcomes, the following equation was used to quantify the transfer effect: ERP(AV) − [ERP(A) + ERP(V)]. This equation subtracted the summed ERPs on unimodal visual and auditory trials from the ERPs on audiovisual trials at each time bin for each electrode [14]. Then, the amplitude of the difference [AV - (A+V)] were compared with 0 at each time point from 0 to 500 ms, using one-sample t tests. If more than 12 consecutive time points were significantly different from zero ($\alpha < 0.05$), audiovisual integration was considered to have occurred. This criterion ensures the reliability of results when a large number of t tests are conducted [135]. Based on the results of the t test, three integration time intervals (80-120 ms, 170-210 ms, and 350-390 ms) and three ROIs (frontal: Fz, F3, F4; central: C3, C4, Cz; and parietal: P3, P4, Pz) were selected for further analysis. In addition, the amplitude of ERPs was averaged across electrodes within each ROI. Repeated-measures ANOVAs with session (pretest, posttest) × ROIs (frontal: Fz, F3, F4; central: C3, C4, Cz; and parietal: P3, P4, Pz) × group (training group, control group) were conducted. The dependent variable was the amplitude difference of [AV - (A+V)].

4.3 Behavioral Results

4.3.1 Training outcomes

For young adults, the ANOVAs based on accuracy revealed a significant main effect of session [$F(1, 35) = 58.299$, $p < 0.001$, $\eta_p^2 = 0.625$], suggesting that accuracy (*d'*) at posttest was increased compared to that at pretest. Significant main effect of load was also found [$F(2, 70) = 80.937$, $p < 0.001$, $\eta_p^2 = 0.698$], indicating that the accuracy decreased with increased working memory load (3-back > 2-back > 1-back). Moreover, a two-way interaction between session and load was found [$F(2, 70) = 7.019$, $p = 0.002$, $\eta_p^2 = 0.167$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction indicated that participants were more accurate at posttest than at pretest across all working memory loads, *ps* < 0.010. Additionally, a significant interaction between load and group was observed [$F(2, 70) = 7.679$, $p = 0.001$, $\eta_p^2 = 0.180$]. After post hoc analysis using pairwise comparisons with a Bonferroni correction, we found that accuracy decreased significantly with increased load in both groups, *ps* < 0.010. Importantly, there was a significant interaction between session and group [$F(1, 35) = 37.152$, $p < 0.001$, $\eta_p^2 = 0.515$]. The post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the accuracy was significantly improved in training group after training, $p < 0.001$. This improvement was not found in the control group, $p = 0.314$. There was a significant three-way interaction [$F(2, 70) = 5.270$, $p = 0.007$, $\eta_p^2 = 0.131$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the training group exhibited significant accuracy improvement at posttest compared to pretest across all working memory loads, *ps* < 0.001. However, when comparing the pretest and posttest in the control group across all working memory loads, there was no significant difference, *ps* > 0.05. In addition, although no significant differences were found between groups for all working memory loads at pretest, the training group had significantly better accuracy than the control group in the 2-back and 3-back conditions at posttest, *ps* < 0.01.

For young adults, analyses of reaction time showed that there was a significant main effect of session [$F(1, 35) = 37.077$, $p < 0.001$, $\eta_p^2 = 0.514$], with the reaction time at posttest being faster than that at pretest. The main effect of load was significant [$F(2, 70) = 51.315$, $p < 0.001$, $\eta_p^2 = 0.595$], with reaction time increasing with higher working memory load (1-back < 2-back < 3-back). The significant main effect of group was also observed [$F(1, 35) = 13.848$, $p = 0.001$, $\eta_p^2 = 0.283$], indicating that the reaction time of the training group was faster than that of the control group. Moreover, a significant interaction between session and load was found [$F(2, 70) = 5.373$, $p < 0.01$, $\eta_p^2 = 0.133$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction revealed that the reaction time at posttest decreased significantly compared to that at pretest across all working memory loads, *ps* < 0.05. It was worth noting that an interaction between session and group was found [$F(1, 35) = 6.366$, $p < 0.05$, $\eta_p^2 = 0.154$]. The post hoc analysis using pairwise comparisons with a Bonferroni correction showed that there was no significant difference between the training group and the control group at pretest, $p = 0.080$. However, the reaction time was significantly decreased in the training group at posttest, $p < 0.001$. The load × group interaction was not significant [$F(2, 70) = 0.800$, $p = 0.453$, $\eta_p^2 = 0.022$]. Notably, the three-way interaction among session, load and group was significant [$F(2, 70) = 11.177$, $p < 0.001$, $\eta_p^2 = 0.242$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction demonstrated that the training group showed a significant reaction time improvement across all working memory loads, *ps* < 0.001. Similar results were found in the 1-back and 2-back conditions in the control group, *ps* < 0.01. Moreover, at posttest, the training group showed significantly decreased reaction time for all working memory loads compared to the control group, *ps* < 0.05. However, the two groups did not show a significant difference at pretest, *ps* $>$ 0.05.

For older adults, the ANOVAs based on accuracy revealed a significant main effect of session, [$F(1, 38) = 9.732$, $p < 0.01$, $\eta_p^2 = 0.204$], indicating that the accuracy was greater at posttest compared to pretest. The main effect of load was also significant, [$F(2, 76) = 264.190$, $p < 0.001$, $\eta_p^2 = 0.874$], indicating that the accuracy decreased with increased load. In addition, the main effect of group was also found, [$F(1, 38) = 18.177$, $p < 0.05$, $\eta_p^2 = 0.249$], suggesting that the over accuracy of training group was greater than that of control group. The two-way interaction between session and load was significant, [$F(2, 76) = 5.839$, $p < 0.01$, $\eta_p^2 = 0.133$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the accuracy of posttest was greater than that of pretest under 3-back level only. Moreover, three-way interaction between session, load, group, [$F(2, 76) = 3.323$, $p < 0.05$, $\eta_p^2 = 0.080$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the training group exhibited greater accuracy at posttest than that at pretest under 3-back level. However, when comparing the pretest and posttest in the control group across all working memory loads, there was no significant difference, *ps* > 0.05.

Analyses of reaction time for older adults showed that there was a significant main effect of load [$F(2, 72) = 37.576$, $p < 0.001$, $\eta_p^2 = 0.511$], with the reaction time being faster at 1-back level than 2-back and 3-back level. The interaction between session and load was significant, [$F(2, 72) = 7.102$, $p < 0.05$, $\eta_p^2 = 0.165$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed the training effect on reaction time was only found at 3-back level, indicating that the reaction time of posttest was faster than that of pretest at 3-back level. No other significant main effect or interaction effects were found (*ps* > 0.050).

Table 1. *The comparison of reaction time and accuracy between pretest and posttest in training and control group for older and young adults.*

	Level	Time	OTG	OCG	YTG	YCG
Rt (ms)	1-back	Pre	700±32	722±34	607±27	718±31
		Post	649±26	728±31	503±23	639±27
	2-back	Pre	870±36	844±38	752±43	866±49
		Post	884±45	939±48	567±28	746±31
	3-back	Pre	1013±58	1069±71	888±44	934±50
		Post	872±56	982±59	585±41	889±46
Accuracy (d')	1-back	Pre	4.2±0.17	3.8±0.17	3.6±0.14	4.0±0.16
		Post	4.4±0.16	3.7±0.15	4.4±0.11	4.1±0.12
	2-back	Pre	2.6±0.20	2.0±0.19	2.8±0.19	3.3±0.22
		Post	2.8±0.23	2.4±0.22	4.0±0.16	3.3±0.18
	3-back	Pre	1.3±0.12	1.1±0.12	1.8±0.21	1.9±0.24
		Post	2.3±0.18	1.4±0.17	4.0±0.15	2.2±0.17

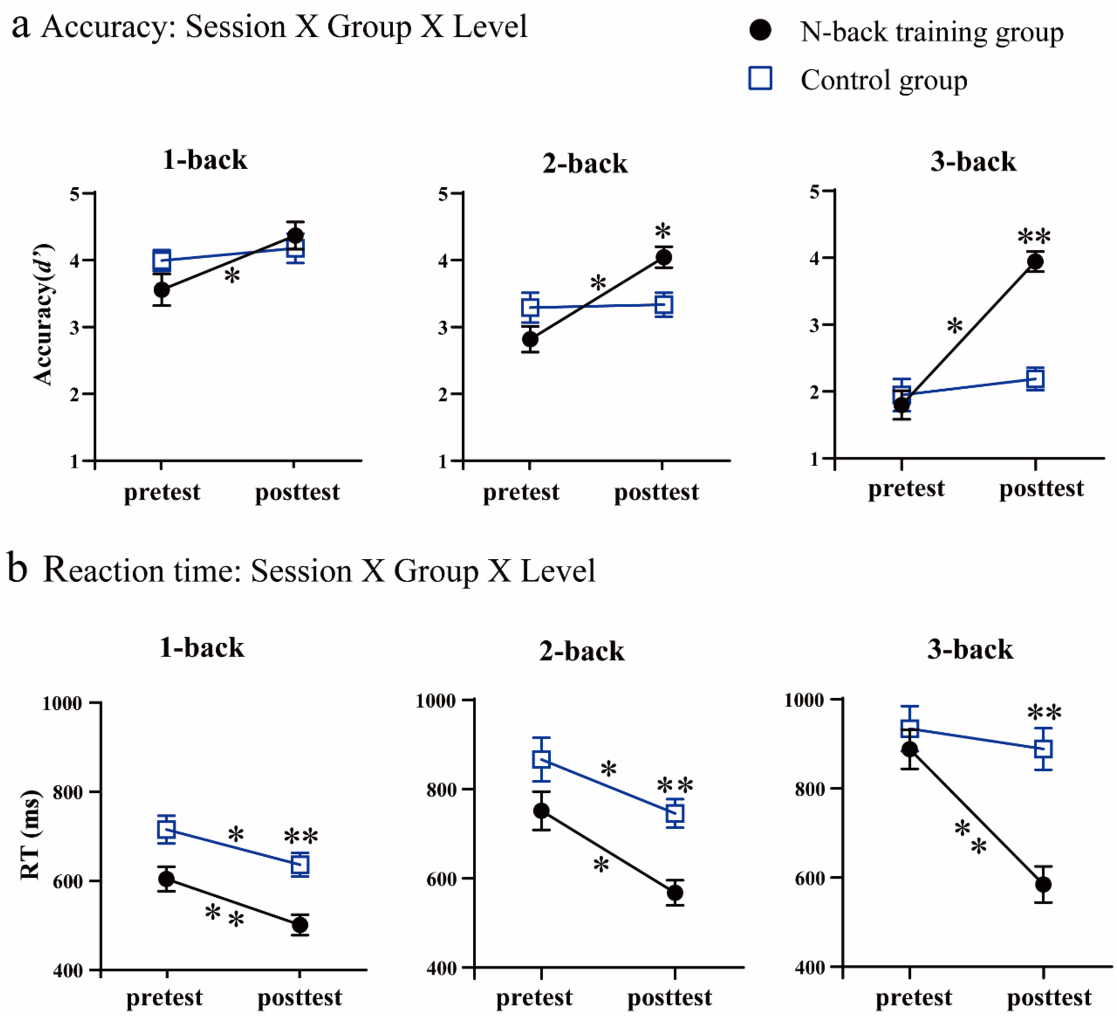
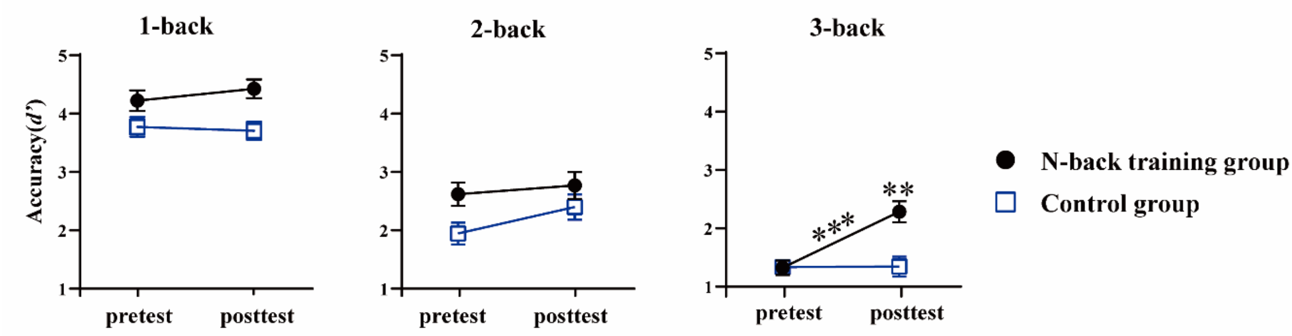


Figure 2. *Changes in performance between pretest and posttest for young adults. (a) Accuracy is depicted for 1-back, 2-back, and 3-back. (b) Reaction times are shown for 1-back, 2-back, and 3-back. The black circles represent the training group, and the blue squares represent the control group. Error bars demonstrated the standard error of the mean (SEM). The asterisk along a trendline indicates the difference between the pretest and posttest, and the asterisk above the black circle or blue squares indicates the difference between different groups at the pretest or posttest, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

c. Accuracy: Session X Group X Level



d. Reaction time: Session X Group X Level

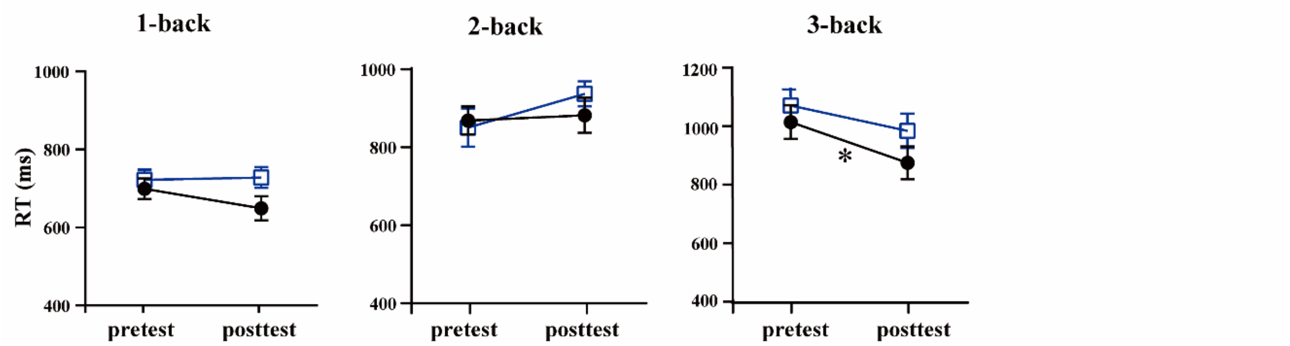


Figure 3. *Changes in performance between pretest and posttest for older adults. (c) Accuracy is depicted for 1-back, 2-back, and 3-back. (d) Reaction times are shown for 1-back, 2-back, and 3-back. The black circles represent the training group, and the blue squares represent the control group. Error bars demonstrated the standard error of the mean (SEM). The asterisk along a trendline indicates the difference between the pretest and posttest, and the asterisk above the black circle or blue squares indicates the difference between different groups at the pretest or posttest, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

4.3.2 Training gain

The ANOVA for training gain in accuracy (*d'*) showed that the main effect of load was significant [$F(2, 70) = 7.01$, $p < 0.01$, $\eta_p^2 = 0.17$], indicating that the training gain in the 3-back condition was greater than that in the 2-back and 1-back conditions. The main effect of group was also significant [$F(1, 35) = 37.20$, $p < 0.001$, $\eta_p^2 = 0.52$], with greater training gain in the training group than in the control group. The interaction between load and group was significant [$F(2, 70) = 5.28$, $p < 0.01$, $\eta_p^2 = 0.13$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the training gain of the training group in the 3-back condition was greater than that in the 2-back and 1-back conditions, *ps* < 0.01. However, the training gain of the control group across the 1-back, 2-back, and 3-back conditions was not significantly different, *ps* > 0.050. In terms of load, greater training gain was found in the training group than in the control group across all working memory levels, *ps* < 0.05.

The ANOVA for training gain in reaction time demonstrated that the main effect of load was significant [$F(2, 70) = 5.373$, $p = 0.07$, $\eta_p^2 = 0.133$], suggesting that the training gain in the 2-back and 3-back conditions was significantly greater than that in the 1-back condition. The main effect of group was also significant [$F(1, 35) = 6.366$, $p = 0.016$, $\eta_p^2 = 0.154$], with greater training gain in the training group than in the control group. The interaction between load and group was significant [$F(2, 70) = 11.177$, $p < 0.001$, $\eta_p^2 = 0.242$]. Further post hoc analysis using pairwise comparisons with a Bonferroni correction indicated that the training gain of the training group in the 3-back condition was greater than that in the 2-back and 1-back conditions, *ps* < 0.050. Similar to the results of training gain in accuracy, there was no difference in training gain across the 1-back, 2-back, and 3-back conditions in the control group, *ps* > 0.050. In terms of load, a greater training gain in the training group was found only in the 3-back condition compared to the control group, $p < 0.001$. In the 1-back and 2-back conditions, no significant difference between two groups was found, *ps* > 0.050.

For older adults, The ANOVA for training gain in accuracy (*d'*) showed that the main effect of load was significant [$F(2, 76) = 5.830$, $p < 0.01$, $\eta_p^2 = 0.133$], indicating that higher training gain was found at 3-back level compared to that at 2-back and 1-back level. Moreover, the interaction between session and group was significant, [F(2, 76) = 3.334, p < 0.01, $\eta_p^2 = 0.081$]. In terms of group, the training gain of training group at 3-back level was greater than that of both 2-back and 1-back level. However, there was no significant difference in training gain across 1-back, 2-back, 3-back level in control group, *ps* > 0.050. In terms of load, significant difference in training gain was found at 3-back level when comparing the training group with control group, indicating that the training gain obtained in training group was greater, *ps* < 0.050.

The ANOVA for training gain in reaction time showed the main effect of load was significant, [F(2, 76) = 5.830, p < 0.01, $\eta_p^2 = 0.133$], suggesting that the training gain was greater at 3-back level than that at 2-back and 3-back level. No other significant main effect or interaction effects were found (*ps* > 0.050).

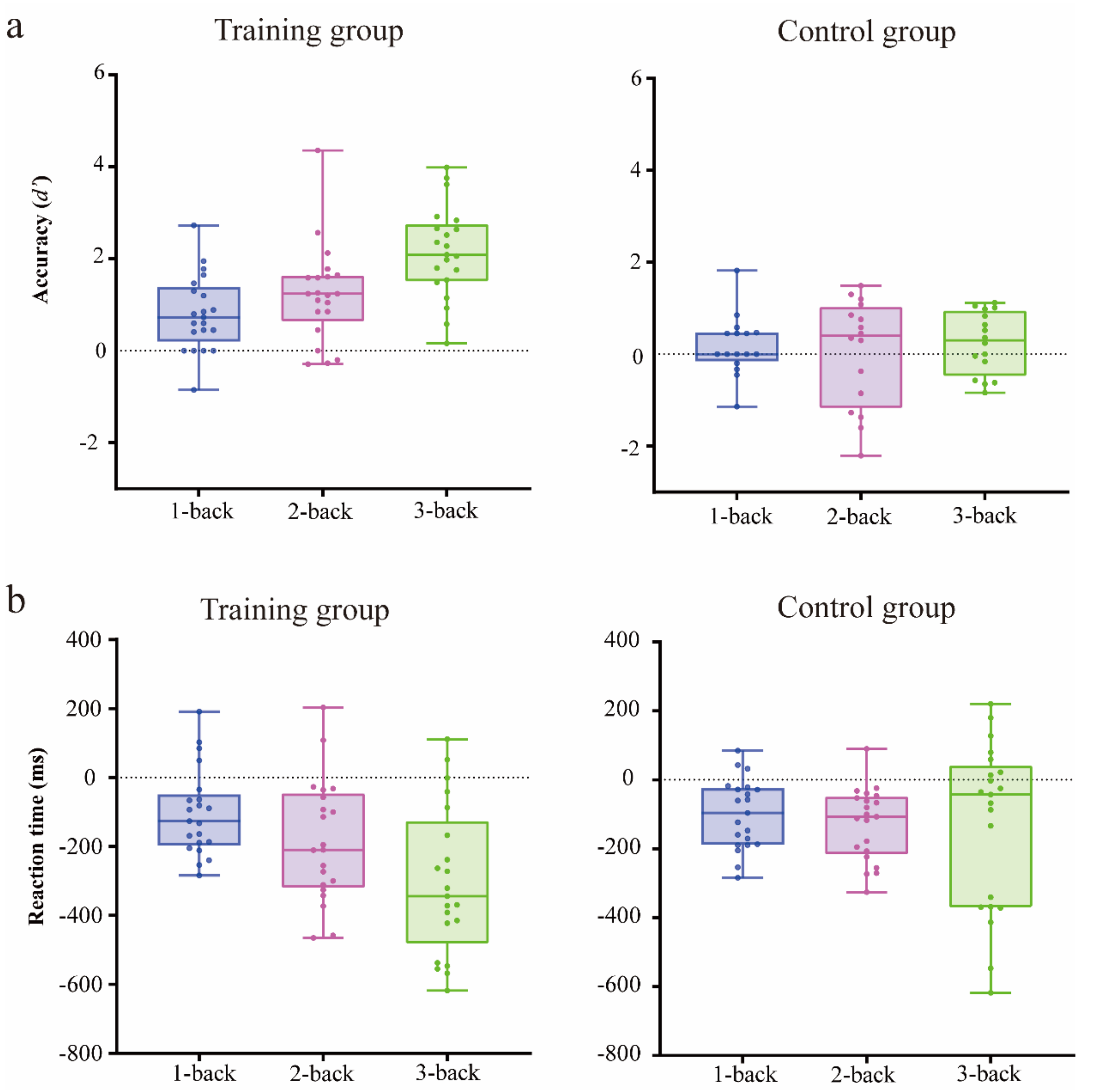


Figure 4. *Box plots of training gain in accuracy (a) and reaction time (b) across 1-back, 2-back, and 3-back in training and control groups for young adults. Median is represented by solid line inside the box represents the median. Upper and lower quartiles are represented by the upper and lower borders respectively.*

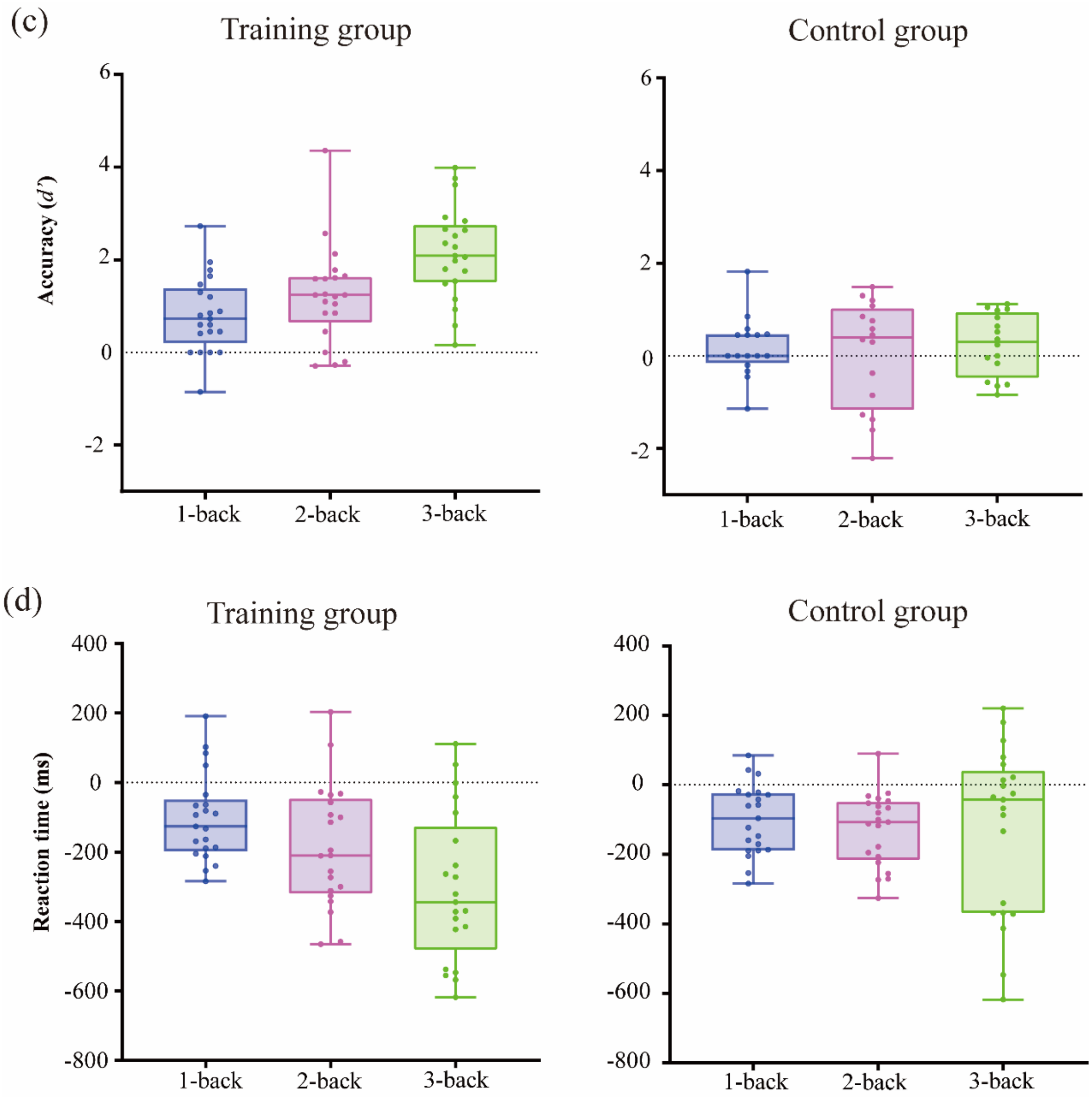


Figure 5. *Box plots of training gain in accuracy (c) and reaction time (d) across 1-back, 2-back, and 3-back in training and control groups for older adults. Median is represented by solid line inside the box represents the median. Upper and lower quartiles are represented by the upper and lower borders respectively.*

4.4 Neural Results

4.4.1 Training outcomes

The N200 amplitude was assessed in the frontal (F3, F4, Fz), central (C3, Cz, C4), and parietal (P3, P4, Pz) regions. In the central region, there was a significant session × group × load interaction [$F(2, 70) = 5.133$, $p < 0.008$, $\eta_p^2 = 0.135$]. Post hoc analyses using pairwise comparisons with a Bonferroni correction indicated that there was no significant difference in the N200 amplitude at pretest between groups, $p > 0.050$. At posttest, however, the training group had significantly greater N200 amplitudes than the control group in the 1-back condition, $p = 0.004$. Notably, significant improvements at posttest compared to pretest in the 3-back condition were found in the training group, $p = 0.041$. There was no change over time in the control group across all working memory loads, *ps* > 0.050. No other significant main effects or interactions were found (*p*s > 0.050). In the central and parietal regions, no significant main effects or interactions were found (*p*s > 0.050).

Then, the P300 amplitude was assessed in these regions. In the frontal regions, a significant session × group interaction was found [$F(1, 35) = 4.228$, $p = 0.048$, $\eta_p^2 = 0.117$]. Post hoc analyses using pairwise comparisons with a Bonferroni correction indicated a pronounced P300 amplitude enhancement at posttest compared to pretest in the training group, $p = 0.053$. However, a significant difference between posttest and pretest in control group was not found, $p > 0.050$. In terms of session, the training group showed no significant difference in P300 amplitude compared to the control group at pretest, $p > 0.050$; similarly, at posttest, there was not a significantly greater P300 amplitude in the training group compared to the control group, $p > 0.050$. No other significant main effects or interactions were found (*p*s > 0.050). In the central regions, there was also a significant session × group interaction [$F(1, 35) = 7.708$, $p = 0.009$, $\eta_p^2 = 0.189$]. Post hoc analyses demonstrated that the training group showed significantly greater P300 amplitudes at posttest than at pretest, $p = 0.008$. There was no significant difference in these values between posttest and pretest in the control group, $p > 0.050$. Moreover, at posttest, the training group showed a significantly greater P300 amplitude than the control group, $p = 0.013$. At pretest, a significant difference between the training group and control group was not observed, $p > 0.050$. No other significant main effects or interactions were found (ps > 0.05). For the parietal region, no significant main effects or interactions were found (ps > 0.05).

For older adults, The N200 amplitude was assessed in the frontal (F3, F4, Fz), central (C3, Cz, C4), and parietal (P3, P4, Pz) regions. In the frontal region, the main effect of sessions was found, [$F(1, 36) = 5.070$, $p = 0.031$, $\eta_p^2 = 0.123$], indicating that the N2 amplitude was greater at posttest than that at pretest. No other significant main effects or interactions were found (ps > 0.05). Although the interaction between session, group, load was not found. The post-hoc analysis showed that the N2 amplitude of training group was greater at posttest than that at pretest under 2-back level only. In the central region and parietal, no significant main effects or interactions were found (ps > 0.05).

In addition, the P300 amplitude was also assessed in the frontal (F3, F4, Fz), central (C3, Cz, C4), and parietal (P3, P4, Pz) regions. However, any significant main effect or interactions were not found (ps > 0.05). Therefore, the N200 latency and P300 latency was assessed in older adults. In the frontal region, the ANONA based on N200 latency was conducted firstly. The main effect of session was found [$F(1, 36) = 4.643$, $p = 0.038$, $\eta_p^2 = 0.114$], indicating that the N200 latency at posttest was earlier than that at pretest. The main effect of load was significant, [$F(2, 72) = 13.664$, $p = 0.000$, $\eta_p^2 =$

0.275], suggesting that the latency of 3-back was advanced compared to that of 2-back and 1-back. Notably, the interaction between session, group and load was found, [$F(2, 72) = 1.521$, $p = 0.225$, $\eta_p^2 = 0.041$]. However, after post-hoc analysis, the significant difference of N200 latency between pretest and posttest at 2- and 3-back levels in training group was found. That is, the N200 latency was advanced at 2- and 3-back levels after training. For the control group, the significant difference of N200 latency between pretest and posttest across 1-, 2-, and 3-back levels were not found. In the central region and partial regions, only main effect of load was found, [$F(2, 72) = 15.915$, $p = 0.000$, $\eta_p^2 = 0.307$], suggesting that the N200 latency was earlier at 3-back level than that at 2- and 1-back levels in central region and the N200 latency was earlier at 3-back level than that at 2-back level in parietal region. In terms of P300 latency, no significant main effects or interactions were found across frontal, central and partial regions (*p*s > 0.05).

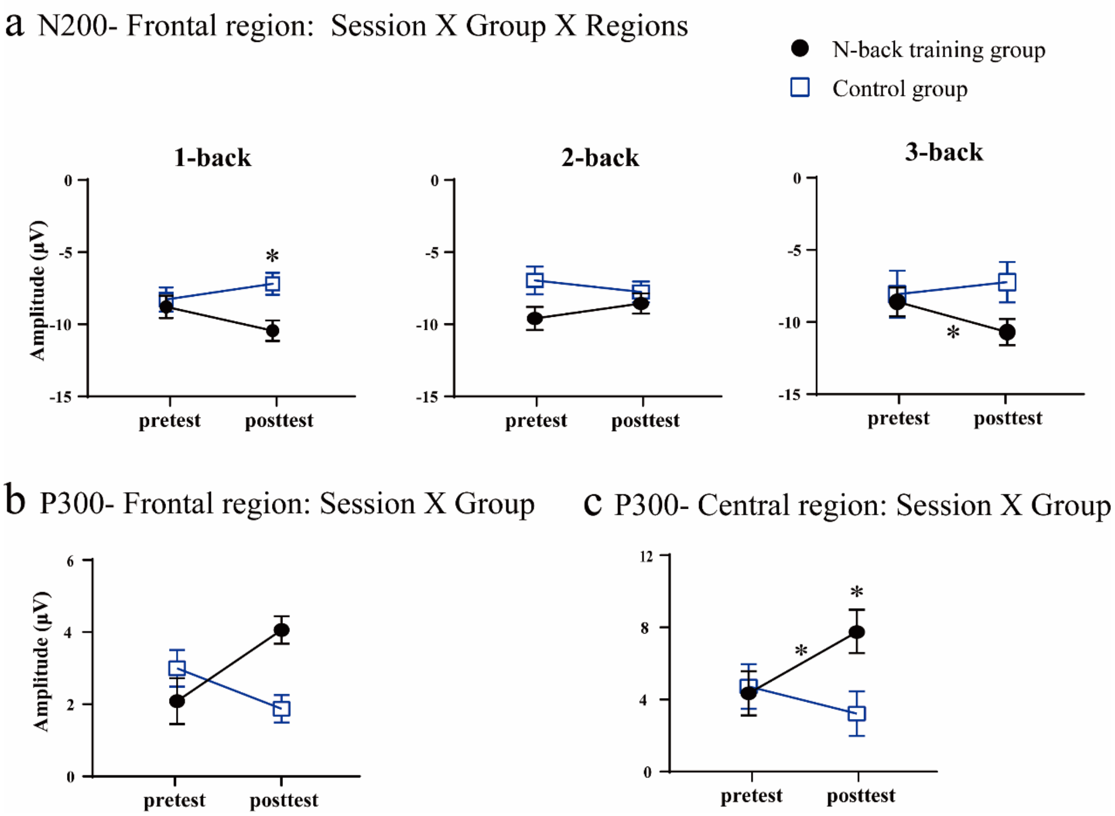


Figure 6. *Training-related effects on ERP components in young adults. Changes in the N200 amplitude in the central region (a) and the P300 amplitude in the frontal region (b), and central region (c) are depicted. Error bars show the SEM. The asterisk along a trendline indicates the difference between the pretest and posttest, and the asterisk above the black circle or blue squares indicates the difference between different groups at the pretest or posttest, * p < 0.05, ** p < 0.01, *** p < 0.001.*

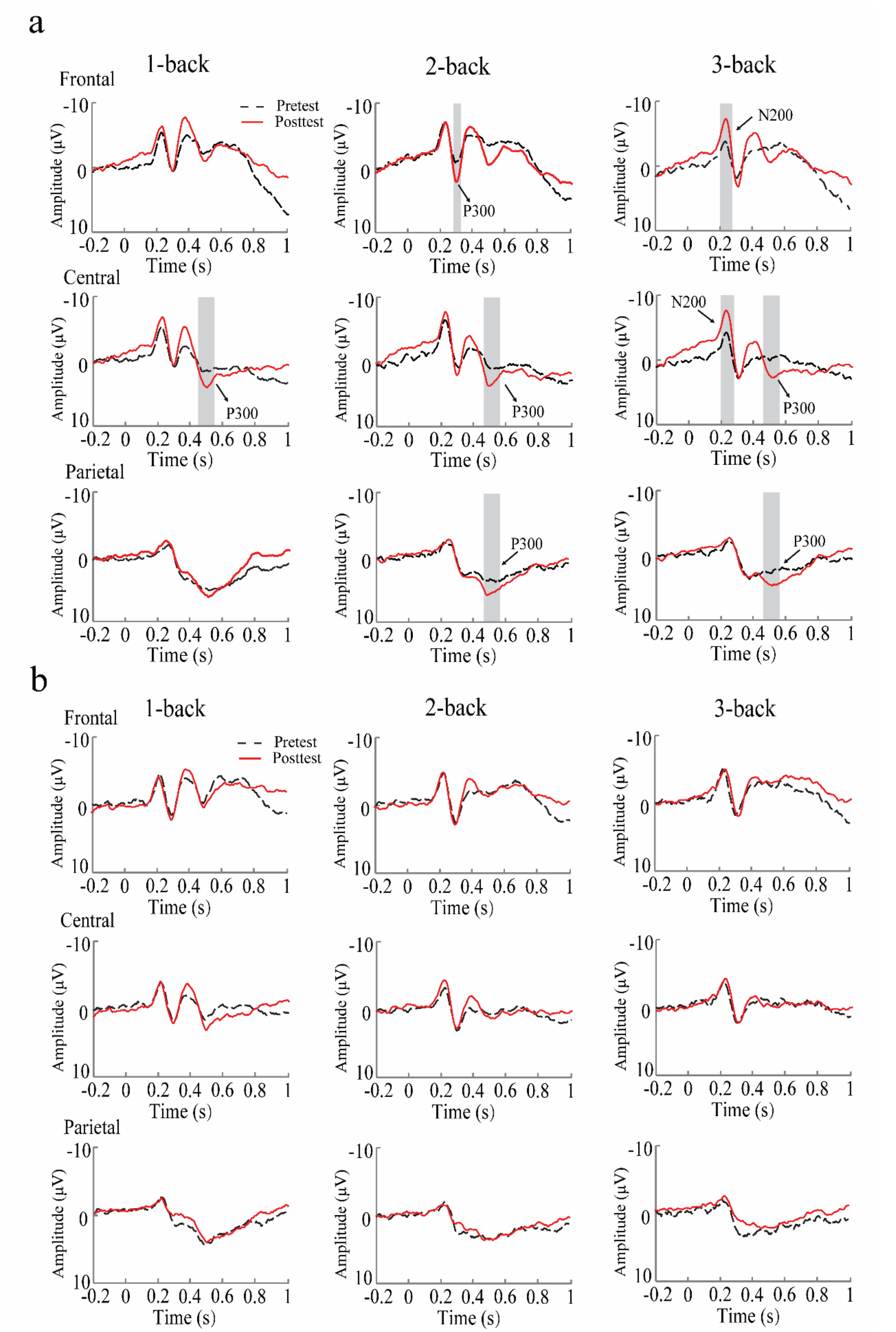


Figure 7. *Event-related potentials (ERPs) at pretest and posttest in the young training group (a) and young control group (b). The gray rectangles highlight marked differences between pretest and posttest.*

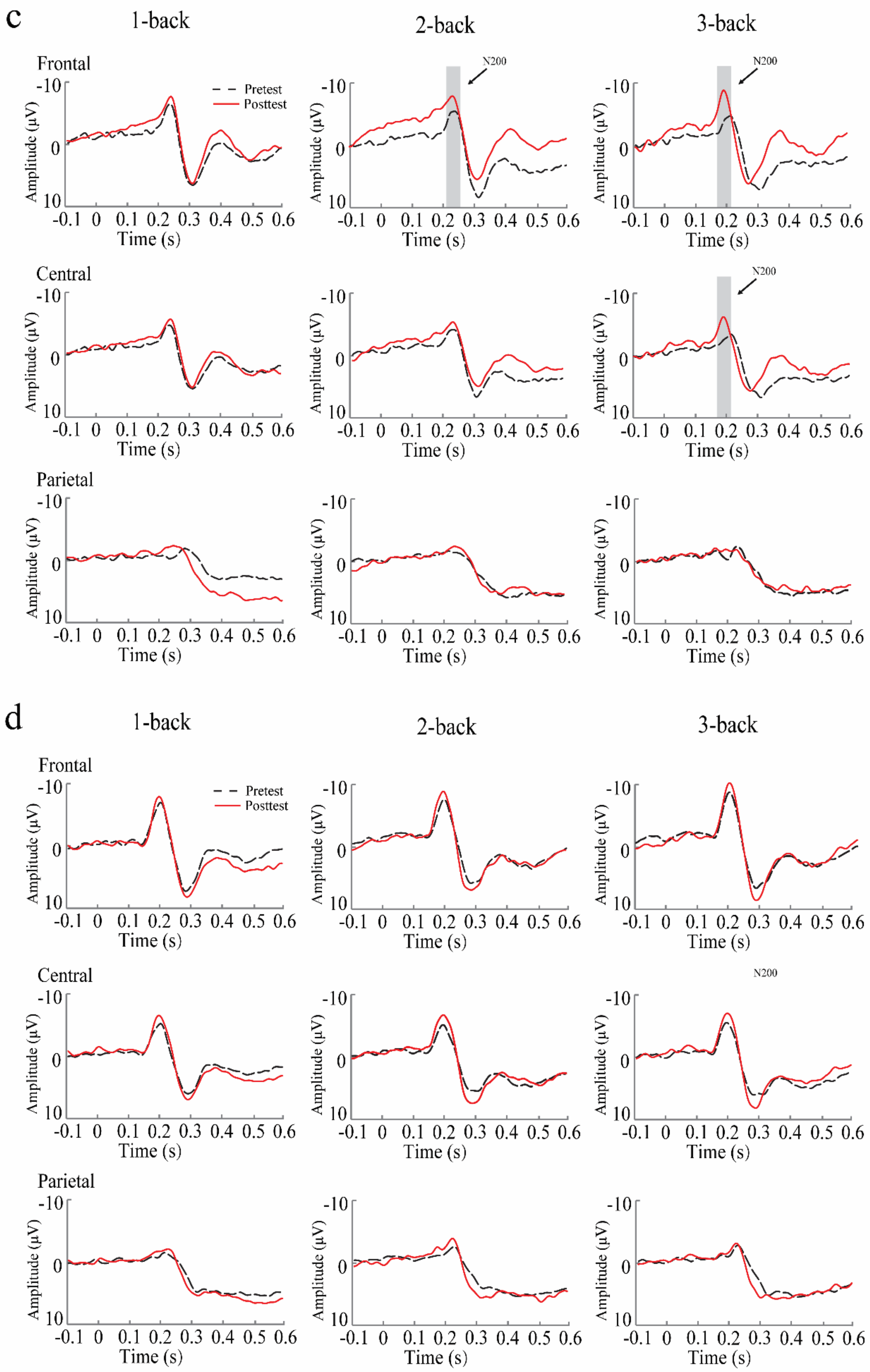


Figure 8. *Event-related potentials (ERPs) at pretest and posttest in the older training group **(c)** and older control group **(d)**. The gray rectangles highlight marked differences between pretest and posttest.*

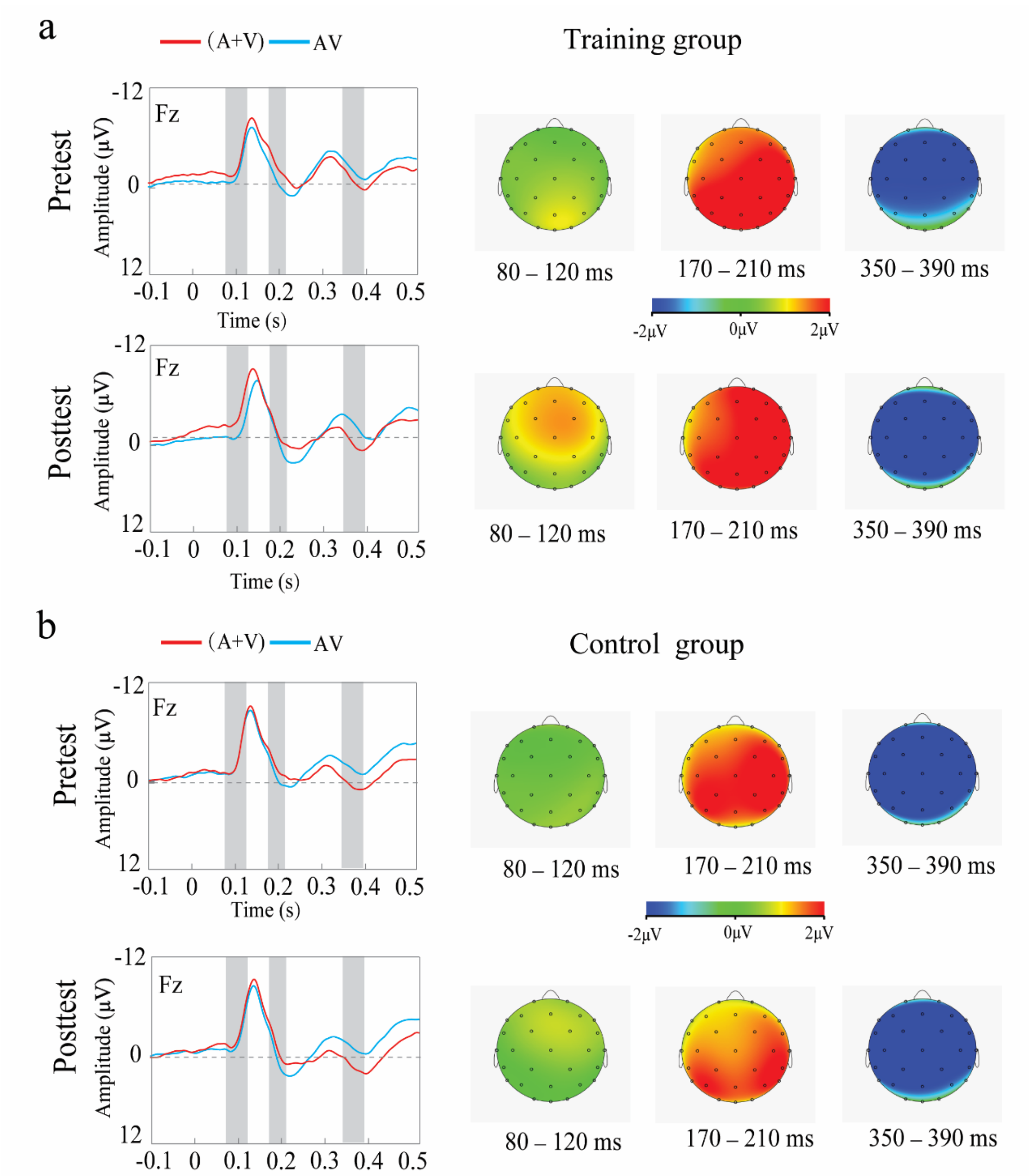


Figure 9. *Grand-average event-related potentials of AV, A+V and the topography map of audiovisual integration in the young training group (a) and control group (b) from 100 ms before stimulus onset to 500 ms after stimulus onset at pretest and posttest.*

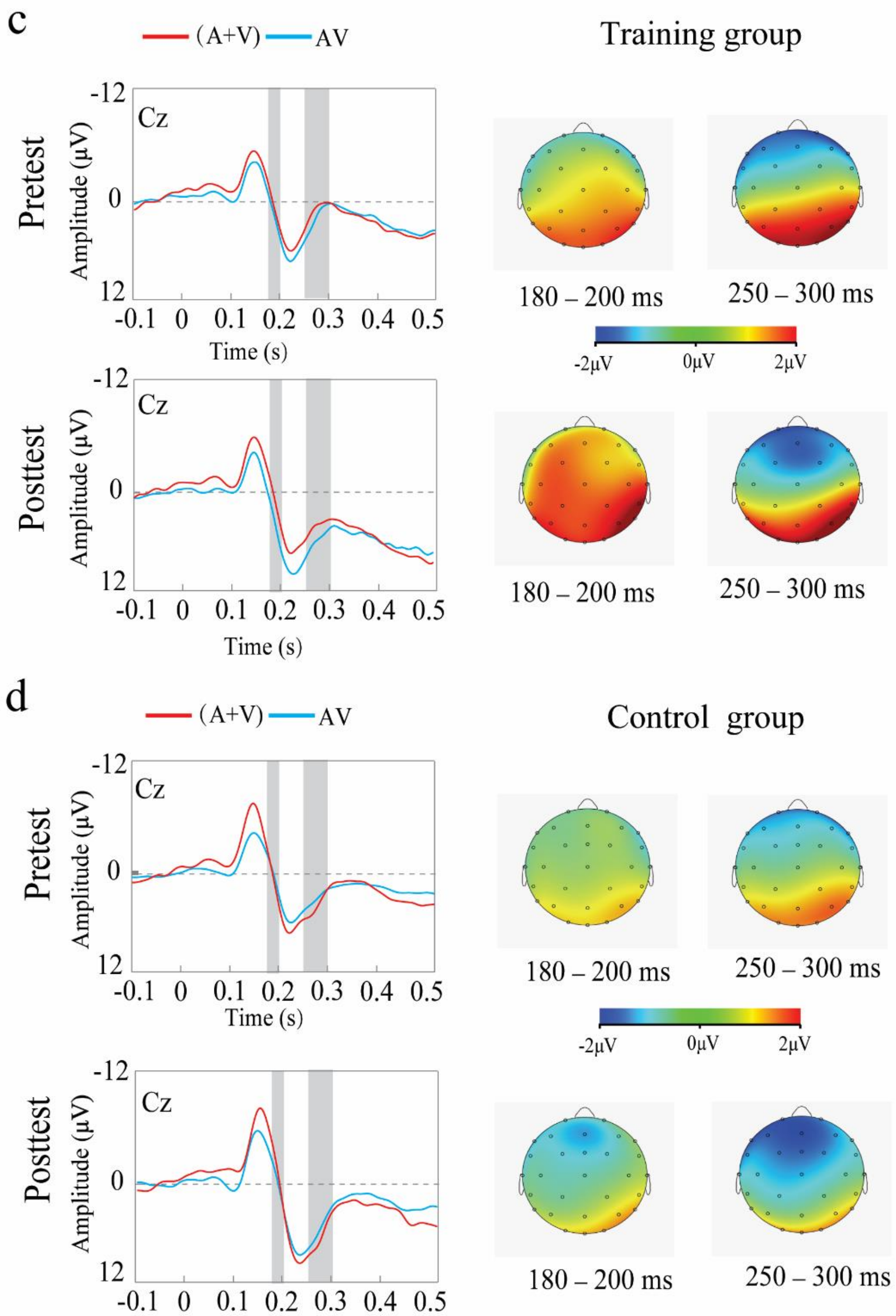


Figure 10. *Grand-average event-related potentials of AV, A+V and the topography map of audiovisual integration in the older training group* ***(c)*** *and control group* ***(d)*** *from 100 ms before stimulus onset to 500 ms after stimulus onset at pretest and posttest.*

4.4.2 Time frequency analysis

For young adults, repeated-measures ANOVAs with sessions (pretest, posttest) × group (training, control) × load (1-back, 2-back, 3-back), were conducted separately to examine alpha neural oscillations in each electrode (Fz, Cz, Pz). At electrode Cz within the time window of 250 - 300 ms. The three-way interaction was found, $F(2, 66) = 3.195$, $p = 0.047$, $\eta_p^2 = 0.088$. The post-hoc analysis showed that the alpha oscillation at 1-back was significantly greater after training for the training group. However, such effect was not found in control group. No main effects or interaction were found in Fz, Pz ($ps > 0.050$). The results suggested that the training could induce the enhancement of alpha oscillations. In addition, the theta oscillation was also analyzed. The results showed that the main effect of sessions was significant, $F(1, 33) = 4.425$, $p = 0.043$, $\eta_p^2 = 0.118$, with post-test neural oscillations being significantly larger than pre-test oscillations. The three-way interaction was found, $F(2, 66) = 3.473$, $p = 0.044$, $\eta_p^2 = 0.095$. After post-hoc analysis, significant theta enhancement was found at 1-back, 3-back in training group, indicating that the training could also induce the alteration of theta. No main effects or interactions were found in Fz, Pz ($ps > 0.050$).

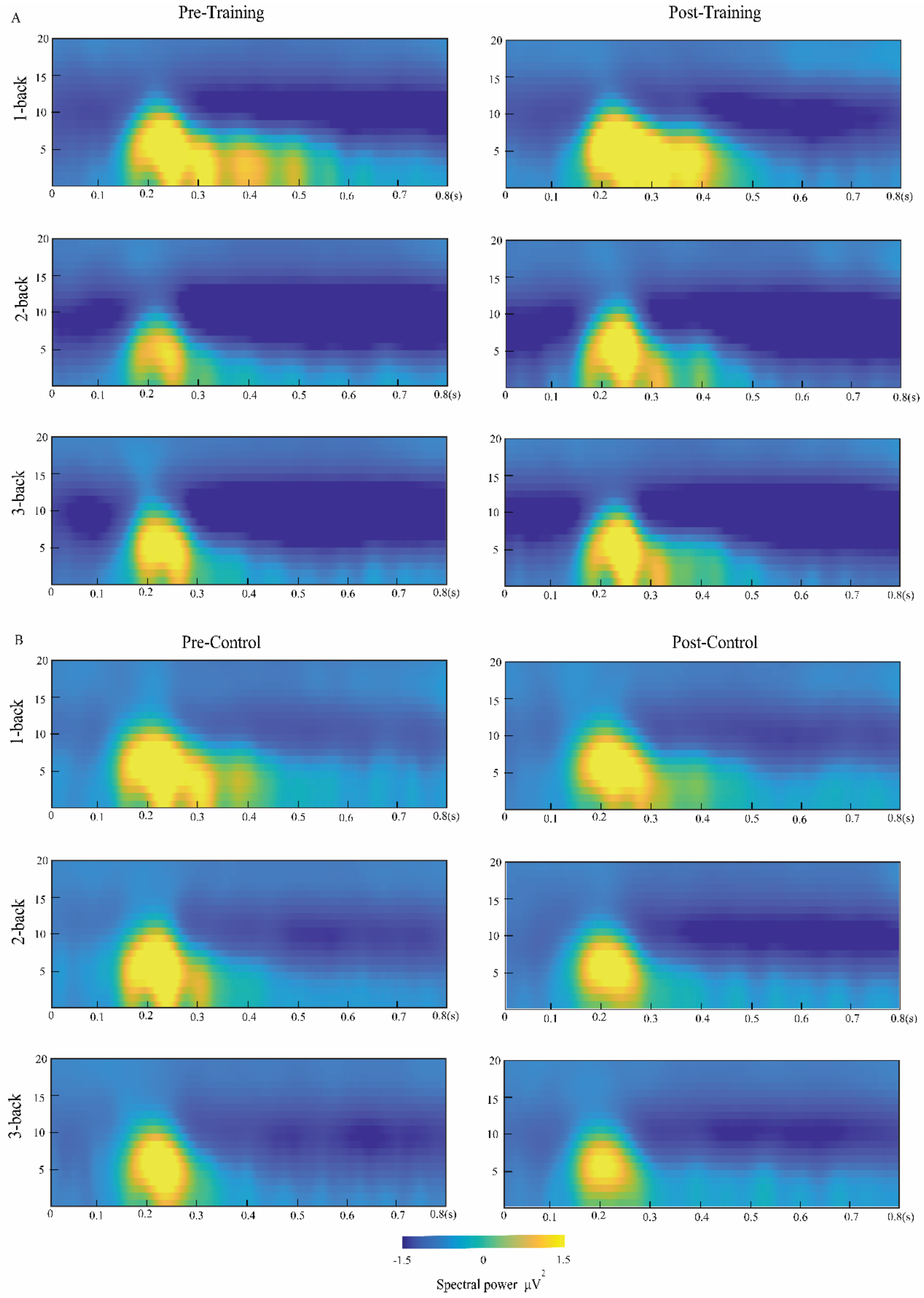


Figure 11. *For the young adults, spectral power activity of 1-back, 2-back, 3-back from 0 - 800 ms across both training group and control group at pretest and posttest.*

For older adults, repeated-measures ANOVAs were conducted separately to examine alpha neural oscillations in each electrode (Fz, Cz, Pz). The content of the analysis included sessions (pretest, posttest) × group (training, control) × load (1-back, 2-back, 3-back). Significant main effect of load was found, $F(2, 56) = 8.56$, $p = 0.01$, $\eta_p^2 = 0.234$, indicating that the alpha oscillations was greater at 1-back and 3-back than at 2-back. The interaction between load and group was also found, $F(2, 56) = 8.56$, $p = 0.001$, $\eta_p^2 = 0.260$. Post-hoc analysis showed that at training group, the alpha oscillation was greater at 1-back compared to 2-back and 3-back. However, at control group, there was no significant difference in the alpha oscillation across 1-back, 2-back, 3-back. Importantly, the three-way interaction between sessions, group and load was found, $F(2, 56) = 8.51$, $p = 0.001$, $\eta_p^2 = 0.233$. Post-hoc analysis showed that for training group, the alpha oscillations were significantly greater at posttest than pretest at 1-back and 2-back. For control group, significant difference in alpha oscillations was not found between pretest and posttest across 1-back, 2-back, 3-back. In the Fz and Cz, no significant main effects or interactions were found (ps > 0.050). In terms of theta oscillations, any main effects or interactions were not found across Fz, Cz, Pz (ps > 0.050).

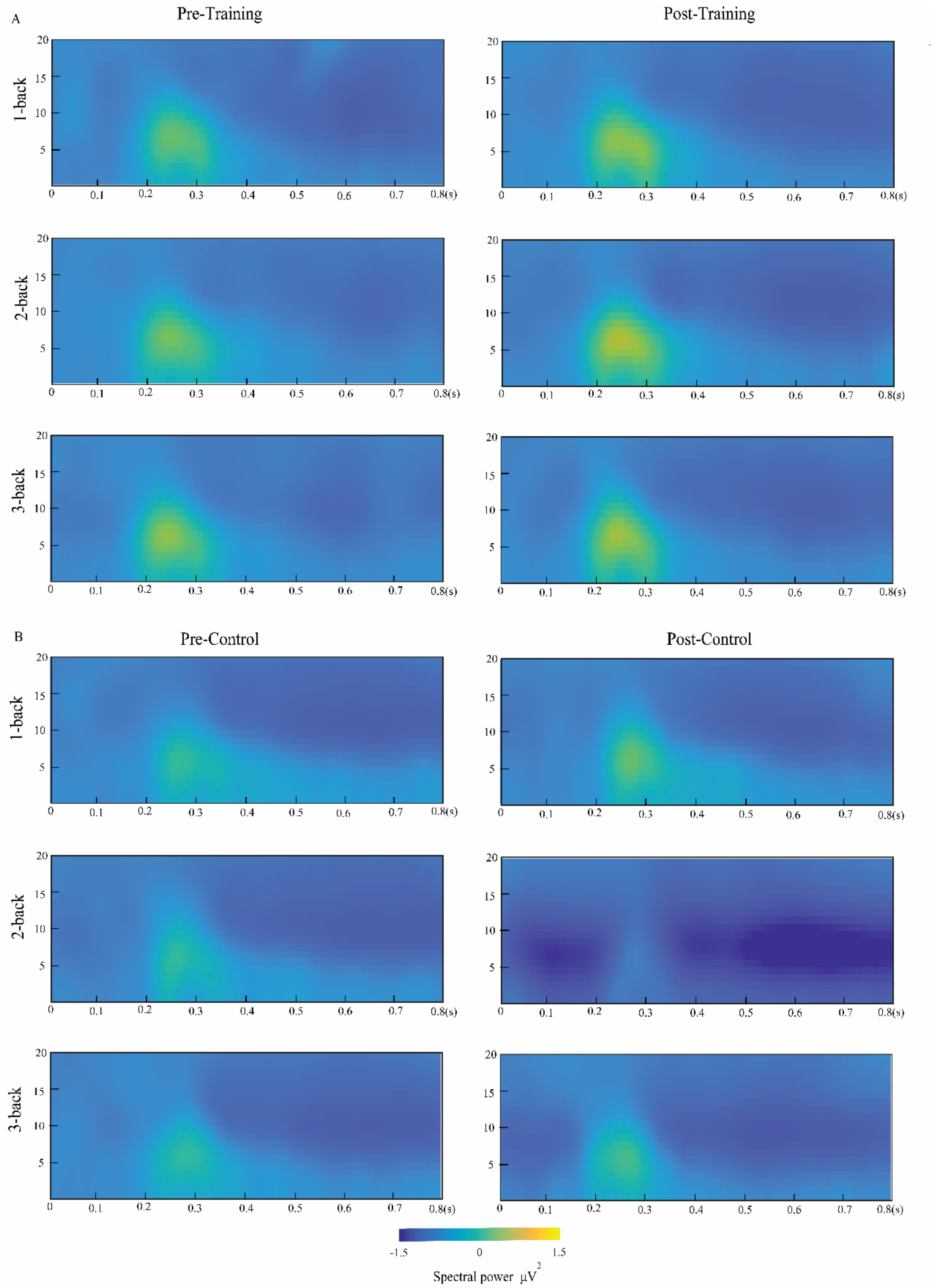


Figure 12. *For the older adults, spectral power activity of 1-back, 2-back, 3-back from 0 - 800 ms across both training group and control group at pretest and posttest.*

4.4.3 Transfer outcomes

At 80-120 ms, the ANOVA showed that the interaction between session and group was significant [$F(1, 35) = 6.181$, $p = 0.018$, $\eta_p^2 = 0.147$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the amplitude at posttest was greater than that at pretest in the training group ($p < 0.05$), whereas such enhancement was not found in control group ($p > 0.05$). In terms of session, the two groups showed no difference at pretest ($p > 0.050$), but a greater amplitude was observed in the training group at posttest ($p = 0.026$). Moreover, a significant three-way interaction was found [$F(2, 70) = 5.245$, $p = 0.007$, $\eta_p^2 = 0.127$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction suggested that there was a significant transfer effect in the frontal region. That is, the amplitude at posttest was greater than that at pretest in the training group, $p = 0.037$. Such a transfer effect was not found in the control group in the frontal region, $p > 0.050$. In the central region, the training group demonstrated a greater amplitude at posttest than at pretest ($p = 0.006$), whereas control group did not find such significant difference ($p > 0.050$). In the parietal region, significant differences between pretest and posttest were not found in either group, *ps* > 0.050. No other significant main effects or interactions were found, *ps* > 0.050. At 170-210 ms, no significant main effects or interactions were found, *ps* > 0.050. At 350-390 ms, the ANOVA showed a significant main effect of region [$F(2, 70) = 4.713$, $p = 0.012$, $\eta_p^2 = 0.116$], indicating that the amplitudes in the frontal and central regions were significantly greater than that in the parietal region. However, no other significant main effects or interactions were found, *ps* > 0.05.

For older adults, at 180-200 ms, the ANOVA showed that the main effect of regions was significant, [$F(3, 108) = 3.913$, $p = 0.011$ $\eta_p^2 = 0.098$], suggesting that amplitude of parietal region was greater than in central, central-parietal regions. The interaction between session and group was significant, [$F(1, 36) = 4.670$, $p = 0.037$ $\eta_p^2 = 0.115$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the amplitude at posttest was greater than that at pretest in the training group ($p < 0.05$), whereas such enhancement was not observed in control group ($p > 0.05$). Moreover, in terms of session, there was no significant differences in pretest between training and control group, whereas significant differences in posttest between training and control group was found. No other significant main effects or interactions were found, *ps* > 0.050. At 250-300 ms, the main effect of region was also significant, [$F(3, 108) = 28.961$, $p = 0.000$ $\eta_p^2 = 0.446$], indicating that the amplitude of occipital regions was the greatest than the other three regions (central, central-parietal, parietal). The interaction between session and group was significant, [$F(1, 36) = 4.207$, $p = 0.048$ $\eta_p^2 = 0.105$]. Post hoc analysis using pairwise comparisons with a Bonferroni correction showed that the amplitude at posttest was greater than that at pretest in the control group ($p < 0.05$), whereas such difference was not observed in training group ($p > 0.05$). No other significant main effects or interactions were found, *ps* > 0.05.

4.5 Discussions

The current study investigated whether audiovisual n-back training induced a training effect on working memory performance and a transfer effect on audiovisual processing. Regarding the training effect, the behavioral results for both age groups showed that training led to improved working memory performance with increased accuracy and decreased reaction time. Moreover, the training group of young adults exhibited greater training gain in the 3-back condition compared to the 1-back and 2-back conditions. For young adults, ERPs analysis found that audiovisual n-back training led to enhancement of the N200 amplitude in the 3-back condition over the frontal area and a higher P300 amplitude in the training group over the frontal and central areas. For older adults, the N200 latency was advanced after training. Regarding the transfer effect, the behavioral results showed that training did not induce significant differences in accuracy and reaction time on the audiovisual processing task between pretest and posttest. However, the ERP results showed that audiovisual integration in the frontal and central regions at 80-120 ms was enhanced in the training group, whereas such an effect was not found in the control group in young adults. For older adults, the audiovisual integration at 180-120ms was enhanced after training in older adults.

4.5.1 Training effect

The training effects appeared as higher accuracy and faster RT at posttest across the 1-back, 2-back, and 3-back conditions in the training group. These results align with previous studies that reported higher accuracy and faster RT after working memory training [71,73,136]. Moreover, we found that the training gain was significantly greater in the 3-back condition and 2-back condition than in the 1-back condition. The available explanation is that higher cognitive load may lead to greater audiovisual benefit [117]. Some researchers have supported this idea by examining whether audiovisual integration leads to better working memory. Their results indicated improved accuracy on the most demanding task (2- and 3-back conditions) for audiovisual stimuli compared to visual or auditory stimuli, whereas such improvement was not found in the 1-back condition. This finding suggests that unimodal conditions may be sufficient for working memory processing when the load is low, but that the advantage of audiovisual benefit becomes obvious when the load increases and more resources are needed to complete the task [87]. Therefore, we infer that greater training effects in the 2-back and 3-back conditions were related to greater audiovisual benefits under higher cognitive loads. Notably, we recruited college students who may possess higher cognitive ability than the general population [91]. In the 1-back condition, a ceiling effect may exist, resulting in no increases in training gain in this condition.

Our results showed that audiovisual working memory training induced a greater amplitude of the P300, consistent with previous findings on working memory training in a single modality [70]. The P300 is a well-established index that reflects working memory updating [70,126]. According to context updating theory, the updating process monitors incoming information and replaces old information that is less relevant to the current task with new information that is more relevant to the current task, thus continuously revising the contents of working memory [137]. Increased P300 amplitudes may reflect better updating ability engaged in working memory processing [70]. Neurocognitive models of the n-back task suggest that encoding, maintenance and updating are necessary processes for working memory [115]. Encoding is critical to working memory processes because it significantly affects subsequent memory processes, such as maintenance and updating [124,138]. Some studies have reported faster working memory updating in an audiovisual working memory encoding condition compared with visual-only or auditory-only working memory encoding conditions [87]. This suggests that the presentation of audiovisual stimuli may elicit more effective encoding, improving the performance of later working memory processes. Therefore, we infer that working memory training with audiovisual stimuli may also exert a positive effect on later updating processes, contributing to the enhancement of the P300 amplitude.

An enhanced N200 amplitude in the frontal region, especially in the most difficult 3-back condition, was also observed. This ERP component is associated with conflict monitoring and mismatch/mismatch identification [123]. Mismatch identification was initially recognized in a sequential matching task, in which participants were required to judge whether a second stimulus matched or mismatched an initial stimulus [123]. The audiovisual n-back task in the current study involved a similar paradigm in which participants were asked to determine the current audiovisual stimulus matched or mismatched the previously audiovisual stimulus maintained in working memory. The audiovisual n-back task in the current study involved a similar paradigm in that participants needed to determine whether there was a mismatch between a currently presented stimulus and a previously presented stimulus maintained in working memory. According to memory trace theory, the presentation of audiovisual stimuli leads to a strong audiovisual memory trace during encoding, and the maintenance of working memory is facilitated due to the factor that it also actives audiovisual memory representation [139]. Because mismatch identification requires a comparison between the currently presented stimulus and the stimulus maintained in working memory, the facilitation of working memory maintenance may increase mismatch identification, reflected by enhancement of the N200 amplitude. In addition, conflict monitoring has also been involved in working memory processing, especially with lure trials [140]. The lure trial in n-back is the same as a previously presented stimulus but is not the matched trial (e.g., 3-back, the second animal cat in the stimuli stream "dog-cat-cow-cat" is lure trial), which may lead to stimulus familiarity and induce a strong conflict effect [74]. Their results confirmed that greater N200 amplitudes were found in the training group at posttest, indicating that conflict monitoring processes are likely engaged in the n-back task with lure trials [74]. However, the task used in our current training task involved solely audiovisual facilitation and did not include lure trials. Therefore, the enhanced N200 amplitude may not be related to improved conflict monitoring.

Notably, the current study found that only the 3-back condition was affected by training, as the N200 amplitude was significantly increased in the training group compared to the control group, consistent with previous studies showing that training induced significant neural alterations in 3-back condition [75]. A possible explanation is that the general executive control process is engaged in more demanding tasks [141]. Early studies compared neural activation across different visual n-back conditions (0-, 1-, 2- and 3-back) and found that more brain areas were activated as difficulty increased (i.e., increasing n). Moreover, the dorsolateral prefrontal cortex (DLPFC) only became involved in the task only under higher demanding conditions (e.g., 2-back, 3-back) [142]. Because the activation of DLPFC is associated with executive functions, this finding indicated that executive functions is particularly engaged under more demanding conditions [143]. We infer that our training may have improved executive functions, as reflected by the greater N200 amplitude in the 3-back condition after training.

For older adults, the audiovisual n-back training led to the alteration of N200 latency, different from the training results of young adults in which found greater amplitude changes after training. Moreover, Older adults showed training effects only on the 3-back task, whereas young adults demonstrated training effects on the 1-back, 2-back, and 3-back tasks. These results indicated that the training effects in older adults were less pronounced compared to young adults. Previous research has compared the effectiveness of training and transfer tasks between young and older adults, consistently finding that young adults exhibit superior performance both before and after training [77]. Additionally, the extent of improvement resulting from training has been observed to be greater in young adults across both trained and untrained tasks. These disparities have been linked to age-related decline in brain structure and function [27], resulting in poor cognitive performance in older adults when their cognitive system are pushed to their limits as in adaptive training. In contrast, young adults appear to exhibit greater adaptability to the increasing task demands, possibly attributable to their possession of a larger pool of neural resources [48]. Consequently, it is expected that young adults will exhibit superior training gains and transfer effects compared to older adults in the present study.

Regarding neural oscillation, we observed an enhancement in both alpha and theta oscillations following the training. These findings consistent with previous studies also reported enhanced neural oscillations [144]. The findings indicated that the training had an impact on the functions associated with maintaining working memory and central executive functions, as evidenced by alterations in alpha and theta oscillations. Concerning theta oscillations, some researchers identified a correlation between theta oscillations and the control mechanisms within working memory. During encoding and retention of information, increased theta oscillation has been observed. Raghavachiari et al., (2001) demonstrated a strong increase of theta activity during the encoding of verbal information in a Stenberg working memory task. Then the author suggested the theta reflected a control mechanism for encoding the sequentially arranged item in working memory [145]. Moreover, some studies explored the relationship between working memory load and theta oscillation. That is, with higher load (higher cognitive processing), the theta oscillation grows stronger [146]. Generally, the index of theta oscillation could be used to measure the alteration of working memory performance. In our research, the audiovisual n-back training task requires participants to monitor the n-back stream and encode the information from each stimulus. The enhancement of the theta activity after training may reflect the improvement of information encoding. For alpha oscillation, some researchers found that alpha oscillation changes were related to the processing of relevant and irrelevant information in working memory [147]. That is, alpha oscillation protected the information that is stored in working memory and inhibited the irrelevant information from the working memory, improving the efficiency of working memory. In our research, the audiovisual n-back training was based on n-back task, which needs to focus on target stimuli (relevant information) and ignore the un-target stimuli (irrelevant information) and requires the updating to new item continuously. We found stronger theta oscillation in young adults after training. This may reflect the improvement of inhibiting ability.

4.5.2 Transfer effect

With respect to the transfer effect, our study showed that audiovisual working memory training induced enhanced audiovisual integration (at 80-120 ms) in the frontal and central regions. Audiovisual integration is interplayed by both early bottom-up and late top-down processing [17]. The amount of available cognitive resources for audiovisual integration may determine the weights of bottom-up and top-down processing [90]. Some evidence is derived from research that investigated how memory load modulates neural oscillations during audiovisual integration. Their results showed that audiovisual integration under high memory load with scarce cognitive resources requires greater top-down processing, reflected by engagement of theta and alpha oscillations [90]. In our research, the transfer task was a simple discrimination task that may have required fewer cognitive resources, suggesting that audiovisual integration was largely governed by bottom-up processing. Moreover, some researchers have explored the latency of bottom-up processing. It appears that the early latency around

100 ms (e.g., N1) reflects a relatively bottom-up process [18]. Therefore, we infer that the enhancement of audiovisual integration at 80-120 ms in our research reflects an improvement in bottom-up processing.

Finally, the current study aimed to investigate the training and transfer effects of audiovisual working memory training, including both behavior and neural outcomes. The results demonstrated that the training led to improved working memory performance as well as enhancement of the N200 and P300 components. Moreover, the training not only successfully improved working memory, but also affected audiovisual processing, reflected by enhanced audiovisual integration at an earlier stage. This implied that higher cognitive functions influenced lower cognitive functions. In summary, these results could give insights into the neural mechanism underlying audiovisual working memory training and lead to a more comprehensive understanding of working memory design.

Chapter 5 Future Projections

Firstly, we demonstrated the facilitating effect of audiovisual integration. Furthermore, we discovered that this facilitating effect of audiovisual integration remains effective in working memory. Lastly, by designing audiovisual working memory training, we found training and transfer effects in both older and young individuals.

Notably, longitudinal studies could be conducted to assess the long-term effects of audiovisual working memory training. Follow-up assessments conducted weeks or months after the training program could help determine the sustainability of the observed improvements and whether they translate into broader cognitive enhancements.

Moreover, considering individual differences, future research could investigate factors such as cognitive abilities, personality traits, and genetic variations that might influence the responsiveness to audiovisual working memory training. Understanding these factors could lead to the development of personalized training approaches tailored to different individuals.

Furthermore, exploring the transfer effects of audiovisual working memory training to other cognitive domains, such as attention or executive functions, would provide valuable insights into the broader impact of this training paradigm.

Finally, the neural mechanisms behind the training effect and transfer effect induced by audiovisual n-back training could be further explored. That is, the future research can investigate the specific neural mechanisms underlying the observed effects of audiovisual working memory training. This could involve using advanced neuroimaging techniques like functional magnetic resonance imaging (fMRI) or magnetoencephalography (MEG) to gain a more detailed understanding of the brain regions and networks involved in working memory processes.

Publications

[1]**Guo, A.,** Yang, W., Yang, X., Lin, J., Li, Z., Ren, Y., . . . Wu, J. (2023). Audiovisual n-Back Training Alters the Neural Processes of Working Memory and Audiovisual Integration: Evidence of Changes in ERPs. *Brain Sciences, 13*(7), 992.

[2]Yang, W., **Guo, A.,** Yao, H., Yang, X., Li, Z., Li, S., . . . Wu, J. (2022). Effect of aging on audiovisual integration: Comparison of high-and low-intensity conditions in a speech discrimination task. *Frontiers in Aging Neuroscience, 14*, 1010060.

[3]Yang, W., Yang, X., **Guo, A.,** Li, S., Li, Z., Lin, J., . . . Zhang, Z. (2022). Audiovisual integration of the dynamic hand-held tool at different stimulus intensities in aging. *Frontiers in Human Neuroscience, 16*, 968987.

[4]Yang, W., Li, S., **Guo, A.,** Li, Z., Yang, X., Ren, Y., . . . Zhang, Z. (2022). Auditory attentional load modulates the temporal dynamics of audiovisual integration in older adults: An ERPs study. *Frontiers in Aging Neuroscience, 14*, 1007954.

Acknowledgement

As I approach the completion of my doctoral degree, I would like to take this opportunity to express my sincere gratitude and heartfelt appreciation to all those who have supported and assisted me along this journey.

Firstly, I would like to extend my deepest thanks to my advisor, Prof. Jinglong Wu. Your expertise, patient guidance, and caring support have been instrumental in my growth during my graduate studies. Your inspiration and mentorship have had a profound impact on me, enabling me to overcome challenges, stay motivated, and strive for excellence. Your encouragement and guidance have been invaluable throughout my entire research endeavor. Beyond his academic contributions, Prof. Jinglong Wu has consistently demonstrated care and support for his students' personal and professional development. His mentorship extends beyond the classroom, and I am grateful for the mentor-mentee relationship we have built.

Secondly, I would like to express special thanks to my master's supervisor, Prof. Weiping Yang. It was under your guidance that I entered the field of cognitive psychology, and step by step, you taught me how to read literature, how to write, and how to design experiments. Over the past six years, you have been a constant source of support for me. Whenever I faced challenges in life or in my studies, you were always there to help me in the first place. Without you, I wouldn't be where I am today. I feel truly fortunate to have had the privilege of being your student. The knowledge, skills, and personal growth I have gained under your guidance will forever remain with me.

As I embark on the next chapter of my career, I will carry the lessons you have taught me and strive to make a positive impact in my field, just as you have done.

Thirdly, I would like to express my gratitude to Prof. Jiajia Yang. Whenever I encountered difficulties during the experiments, you provided me with detailed answers and helped me solve the problems. The smooth progress of the experiments would not have been possible without your support. I am immensely grateful for the opportunities provided by Prof. Jiajia Yang to collaborate on research projects, attend conferences, and engage in discussions.

Fourthly, I would like to express my gratitude to Prof. Satoshi Takahashi. His rigorous academic attitude and meticulous work ethic are truly worthy of my admiration and learning. I have benefited greatly from his precise feedback during our group meetings, which encompass various aspects such as Japanese language communication skills, experimental design, and data visualization.

Fifthly, I also want to express my gratitude to Prof. Yanna Ren. Whenever I encountered programming issues during the experiments, it was you who proactively assisted me in resolving code execution problems and generously shared the functioning of the experimental programs, which greatly benefited me. Furthermore, I sincerely appreciate your tireless efforts in revising my thesis. It is through your modifications and suggestions that the quality of my thesis has significantly improved.

Finally, I am immensely thankful to all the students in the Wu lab for the countless hours we have dedicated to working together in the lab, where we discuss papers, conducted experiments, and analyzed data. Your unwavering enthusiasm, boundless creativity, and steadfast dedication have consistently motivated me to push the limits of my research and pursue excellence. The prevailing spirit of collaboration within our lab has enabled us to collectively tackle intricate challenges and accomplish remarkable outcomes.

References

[1]Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience, 17*(9), 1396-1409. doi:10.1162/0898929054985383

[2]Yang, W., Yang, J., Gao, Y., Tang, X., Ren, Y., Takahashi, S., & Wu, J. (2015). Effects of Sound Frequency on Audiovisual Integration: An Event-Related Potential Study. *PloS one, 10*(9), 1-15. doi:10.1371/journal.pone.0138296

[3]Winneke, A. H., & Phillips, N. A. (2011). Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. *Psychol Aging, 26*(2), 427-438. doi:10.1037/a0021683

[4]Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris, 98*(1-3), 191-205. doi:10.1016/j.jphysparis.2004.03.018

[5]Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neurosci Lett, 442*(3), 257-261. doi:10.1016/j.neulet.2008.07.010

[6]Spence, C. (2007). Audiovisual multisensory integration. *Acoustical science and technology, 28*(2), 61-70.

[7]Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci, 9*(4), 255-266. doi:10.1038/nrn2331

[8]Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology, 10*(11), 649-657.

[9]Bedard, G., & Barnett-Cowan, M. (2016). Impaired timing of audiovisual events in the elderly. *Exp Brain Res, 234*(1), 331-340. doi:10.1007/s00221-015-4466-7

[10]Senkowski, D., Saint-Amour, D., Hofle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage, 56*(4), 2200-2208. doi:10.1016/j.neuroimage.2011.03.075

[11]Yang, W., Guo, A., Yao, H., Yang, X., Li, Z., Li, S., . . . Wu, J. (2022). Effect of aging on audiovisual integration: Comparison of high-and low-intensity conditions in a speech discrimination task. *Frontiers in Aging Neuroscience, 14*, 1010060.

[12]Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage, 44*(3), 1210-1223. doi:10.1016/j.neuroimage.2008.09.034

[13]Freiherr, J., Lundstrom, J. N., Habel, U., & Reetz, K. (2013). Multisensory integration mechanisms during aging. *Front Hum Neurosci, 7*, 863. doi:10.3389/fnhum.2013.00863

[14]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.

[15]Yang, W., Li, Z., Guo, A., Li, S., Yang, X., & Ren, Y. (2021). Effects of stimulus intensity on audiovisual integration in aging across the temporal dynamics of processing. *International*

Journal of Psychophysiology, 162, 95-103.

[16]Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex, 68*, 169-181. doi:10.1016/j.cortex.2015.03.006

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

[18]Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia, 53*, 115-121. doi:10.1016/j.neuropsychologia.2013.11.011

[19]Miller, L. M., & D'esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience, 25*(25), 5884-5893.

[20]Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of Cognitive Neuroscience, 16*(6), 1000-1009.

[21]Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The journal of neuroscience, 32*(25), 8443-8453. doi:10.1523/JNEUROSCI.5069-11.2012

[22]Chan, J. S., Wibral, M., Wollstadt, P., Stawowsky, C., Brandl, M., Helbling, S., . . . Kaiser, J. (2017). Predictive coding over the lifespan: Increased reliance on perceptual priors in older adults–a magnetoencephalography and dynamic causal modelling study. *bioRxiv, 2*(18), 80-95. doi:10.1101/178095

[23]Alm, M., & Behne, D. (2013). Audio-visual speech experience with age influences perceived audio-visual asynchrony in speech. *The Journal of the Acoustical Society of America, 134*(4), 3001-3010. doi:10.1121/1.4820798

[24]Liu, X., & Yan, D. (2007). Ageing and hearing loss. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 211*(2), 188-197. doi:10.1002/path.2102

[25]Bäckman, L., Nyberg, L., Lindenberger, U., Li, S.-C., & Farde, L. (2006). The correlative triad among aging, dopamine, and cognition: current status and future prospects. *Neuroscience & Biobehavioral Reviews, 30*(6), 791-807. doi:10.1016/j.neubiorev.2006.06.005

[26]Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2012). *Multisensory Integration and Aging*: The Neural Bases of Multisensory Processes.

[27]Grady, C. (2012). The cognitive neuroscience of ageing. *Nature reviews neuroscience, 13*(7), 491-505.

[28]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.

[29]Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport, 18*(10), 1077-1081. doi:10.1097/WNR.0b013e3281e72ae7

[30]Diaconescu, A. O., Hasher, L., & McIntosh, A. R. (2013). Visual dominance and multisensory integration changes with age. *Neuroimage, 65*, 152-166.

[31]Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental brain research, 227*, 249-261.

[32]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 Hear, 26*(3), 263-275. doi:10.1097/00003446-200506000-00003

[33]Zou, Z., Chau, B. K. H., Ting, K. H., & Chan, C. C. H. (2017). Aging Effect on Audiovisual Integrative Processing in Spatial Discrimination Task. *Front Aging Neurosci, 9*, 374. doi:10.3389/fnagi.2017.00374

[34]Fabiani, M. (2012). It was the best of times, it was the worst of times: a psychophysiologist's view of cognitive aging. *Psychophysiology, 49*(3), 283-304. doi:10.1111/j.1469-8986.2011.01331.x

[35]Cabeza, R., Grady, C. L., Nyberg, L., McIntosh, A. R., Tulving, E., Kapur, S., . . . Craik, F. I. (1997). Age-related differences in neural activity during memory encoding and retrieval: a positron emission tomography study. *Journal of Neuroscience, 17*(1), 391-400.

[36]Rypma, B., Eldreth, D. A., & Rebbechi, D. (2007). Age-related differences in activation-performance relations in delayed-response tasks: a multiple component analysis. *Cortex, 43*(1), 65-76.

[37]Hugenschmidt, C. E., Mozolic, J. L., & Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport, 20*(4), 349-353. doi:10.1097/WNR.0b013e328323ab07

[38]Hugenschmidt, C. E., Mozolic, J. L., Tan, H., Kraft, R. A., & Laurienti, P. J. (2009). Age-related increase in cross-sensory noise in resting and steady-state cerebral perfusion. *Brain topography, 21*(3), 241-251. doi:10.1007/s10548-009-0098-1

[39]Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults. *Frontiers in Psychology, 5*(323), 1-12. doi:https://doi.org/10.3389/fpsyg.2014.00323

[40]DeLoss, D. J., Pierce, R. S., & Andersen, G. J. (2013). Multisensory integration, aging, and the sound-induced flash illusion. *Psychology and Aging, 28*(3), 802-812. doi: https://doi.org/10.1037/a0033289

[41]Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience, 4*(10), 829-839. doi:10.1038/nrn1201.

[42]Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior, 19*(4), 450-466.

[43]Kaufman, A. S., & Kaufman, N. L. (2015). *Essentials of working memory assessment and intervention*: John Wiley & Sons.

[44]Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 336.

[45]Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of experimental psychology: General, 130*(2), 169.

[46]Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and aging, 17*(1), 85.

[47]Reuter-Lorenz, P. A., Jonides, J., Smith, E. E., Hartley, A., Miller, A., Marshuetz, C., & Koeppe, R. A. (2000). Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of Cognitive Neuroscience, 12*(1), 174-187.

[48]Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current directions in psychological science, 17*(3), 177-182.

[49]Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., . . . Weinberger, D. R. (2006). Neurophysiological correlates of age-related changes in working memory capacity. *Neuroscience Letters, 392*(1-2), 32-37.

[50]Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø., Larsen, V. A., & Walhovd, K. B. (2010). Effects of memory training on cortical thickness in the elderly. *Neuroimage, 52*(4), 1667-1676.

[51]Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., & Bäckman, L. (2011). Neural correlates of training-related working-memory gains in old age. *Neuroimage, 58*(4), 1110-1120.

[52]Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø., Larsen, V. A., & Walhovd, K. B. (2012). Memory training impacts short-term changes in aging white matter: A longitudinal diffusion tensor imaging study. *Human brain mapping, 33*(10), 2390-2406.

[53]Hempel, A., Giesel, F. L., Garcia Caraballo, N. M., Amann, M., Meyer, H., Wüstenberg, T., . . . Schröder, J. (2004). Plasticity of cortical activation related to working memory during training. *American Journal of Psychiatry, 161*(4), 745-747.

[54]Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature neuroscience, 7*(1), 75-79.

[55]Stevens, M. C., Gaynor, A., Bessette, K. L., & Pearlson, G. D. (2016). A preliminary study of the effects of working memory training on brain function. *Brain imaging and behavior, 10*, 387-407.

[56]Kühn, S., Schmiedek, F., Noack, H., Wenger, E., Bodammer, N. C., Lindenberger, U., & Lövden, M. (2013). The dynamics of change in striatal activity following updating training. *Human brain mapping, 34*(7), 1530-1541.

[57]Schneiders, J. A., Opitz, B., Tang, H., Deng, Y., Xie, C., Li, H., & Mecklinger, A. (2012). The impact of auditory working memory training on the fronto-parietal working memory network. *Frontiers in Human Neuroscience, 6*, 173.

[58]Schneiders, J. A., Opitz, B., Krick, C. M., & Mecklinger, A. (2011). Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. *Cerebral cortex, 21*(11), 2555-2564.

[59]Clark, C. M., Lawlor-Savage, L., & Goghari, V. M. (2017). Functional brain activation associated with working memory training and transfer. *Behavioural brain research, 334*, 34-49.

[60]Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & cognition, 42*, 464-480.

[61]Jolles, D. D., van Buchem, M. A., Crone, E. A., & Rombouts, S. A. (2013). Functional brain connectivity at rest changes after working memory training. *Human brain mapping, 34*(2), 396-406.

[62]Li, W., Zhang, Q., Qiao, H., Jin, D., Ngetich, R. K., Zhang, J., . . . Li, L. (2021). Dual n-back working memory training evinces superior transfer effects compared to the method of loci. *Scientific Reports, 11*(1), 3072.

[63]Nyberg, L., Sandblom, J., Jones, S., Neely, A. S., Petersson, K. M., Ingvar, M., & Bäckman, L. (2003). Neural correlates of training-related memory improvement in adulthood and aging. *Proceedings of the National Academy of Sciences, 100*(23), 13728-13733.

[64]Kirchhoff, B., Anderson, B., Barch, D., & Jacoby, L. (2012). Cognitive and neural effects of semantic encoding strategy training in older adults. *Cerebral cortex, 22*(4), 788-799.

[65]Karbach, J., & Kray, J. How useful is executive control training. *Age differences in near and far*.

[66]Kray, J., Karbach, J., Haenig, S., & Freitag, C. (2012). Can task-switching training enhance executive

control functioning in children with attention deficit/-hyperactivity disorder? *Frontiers in Human Neuroscience, 5*, 180.

[67]Johann, V. E., & Karbach, J. (2020). Effects of game-based and standard executive control training on cognitive and academic abilities in elementary school children. *Developmental science, 23*(4), e12866.

[68]Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., . . . Kramer, A. F. (2007). Training-induced functional activation changes in dual-task processing: an FMRI study. *Cerebral cortex, 17*(1), 192-204.

[69]Zhao, X., & Jia, L. (2019). Training and transfer effects of interference control training in children and young adults. *Psychological research, 83*(7), 1519-1530.

[70]Zhao, X., Zhou, R., & Fu, L. (2013). Working memory updating function training influenced brain activity. *PloS one, 8*(8), 1-7. doi:10.1371/journal.pone.0071063.

[71]Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829-6833.

[72]Salminen, T., Frensch, P., Strobach, T., & Schubert, T. (2016). Age-specific differences of dual n-back training. *Aging, Neuropsychology, and Cognition, 23*(1), 18-39.

[73]Küper, K., & Karbach, J. (2016). Increased training complexity reduces the effectiveness of brief working memory training: evidence from short-term single and dual n-back training interventions. *Journal of Cognitive Psychology, 28*(2), 199-208.

[74]Covey, T. J., Shucard, J. L., & Shucard, D. W. (2019). Working memory training and perceptual discrimination training impact overlapping and distinct neurocognitive processes: Evidence from event-related potentials and transfer of training gains. *Cognition, 182*, 50-72. doi:10.1016/j.cognition.2018.08.012

[75]Pergher, V., Wittevrongel, B., Tournoy, J., Schoenmakers, B., & Van Hulle, M. M. (2018). N-back training and transfer effects revealed by behavioral responses and EEG. *Brain and Behavior, 8*(11), e01136.

[76]Xu, J., Deng, M., Nan, W., & Cai, D. (2021). The effects of working memory training in children revealed by behavioral responses and ERP. *Brain and Behavior, 11*(8), e2310.

[77]Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320*(5882), 1510-1512.

[78]Salminen, T., Kuhn, S., Frensch, P. A., & Schubert, T. (2016). Transfer after Dual n-Back Training Depends on Striatal Activation Change. *J Neurosci, 36*(39), 10198-10213. doi:10.1523/JNEUROSCI.2305-15.2016

[79]Miró-Padilla, A., Bueichekú, E., Ventura-Campos, N., Flores-Compañ, M.-J., Parcet, M. A., & Ávila, C. (2019). Long-term brain effects of N-back training: an fMRI study. *Brain imaging and behavior, 13*, 1115-1127.

[80]Grützmann, R., Kathmann, N., Gutmann, G., & Heinzel, S. (2021). Effects of adaptive and non-adaptive three-week executive control training on interference control: Evidence from the N2, CRN, and ERN. *International Journal of Psychophysiology, 162*, 8-21.

[81]Jansen, S. D., Keebler, J. R., & Chaparro, A. (2018). Shifts in maximum audiovisual integration with age. *Multisensory Research, 31*(3-4), 191-212.

[82]Chen, Y.-C., Yeh, S.-L., & Tang, P.-F. (2021). Age-related changes in audiovisual simultaneity

perception and their relationship with working memory. *The Journals of Gerontology: Series B, 76*(6), 1095-1103.

[83]Heikkilä, J., & Tiippana, K. (2016). School-aged children can benefit from audiovisual semantic congruency during memory encoding. *Experimental brain research, 234*, 1199-1207.

[84]Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139*(1), 23-38.

[85]Druzgal, T. J., & D'esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience, 15*(6), 771-784.

[86]Xie, Y. J., Li, Y. Y., Xie, B., Xu, Y. Y., & Peng, L. (2019). The neural basis of complex audiovisual objects maintenances in working memory. *Neuropsychologia, 133*, 107189. doi:10.1016/j.neuropsychologia.2019.107189

[87]Frtusova, J. B., Winneke, A. H., & Phillips, N. A. (2013). ERP evidence that auditory–visual speech facilitates working memory in younger and older adults. *Psychology and aging, 28*(2), 481.

[88]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), 5112.

[89]Xie, Y., Li, Y., Duan, H., Xu, X., Zhang, W., & Fang, P. (2021). Theta oscillations and source connectivity during complex audiovisual object encoding in working memory. *Frontiers in Human Neuroscience, 15*, 614950.

[90]Michail, G., Senkowski, D., Niedeggen, M., & Keil, J. (2021). Memory load alters perception-related neural oscillations during multisensory integration. *Journal of Neuroscience, 41*(7), 1505-1515.

[91]He, Y., Guo, Z., Wang, X., Sun, K., Lin, X., Wang, X., . . . Zhang, J. (2022). Effects of Audiovisual Interactions on Working Memory Task Performance—Interference or Facilitation. *Brain Sciences, 12*(7), 886.

[92]Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS biology, 4*(10), e326.

[93]Deveau, J., Jaeggi, S. M., Zordan, V., Phung, C., & Seitz, A. R. (2015). How to build better memory training games. *Frontiers in systems neuroscience, 8*, 243.

[94]Wallace, M. T. (2004). The Development of Multisensory Integration. *Perception, 26*(1), 76-77.

[95]Spear, P. D. (1993). Neural bases of visual deficits during aging. *Vision Research, 33*(18), 2589-2609.

[96]Chiang-Shan Ray, L., Peisi, Y., Bergquist, K. L., & Rajita, S. (2007). Greater activation of the "default" brain regions predicts stop signal errors. *Neuroimage, 38*(3), 640-648.

[97]Stephen, J. M., Knoefel, J. E., Adair, J., Hart, B., & Aine, C. J. (2010). Aging-related changes in auditory and visual integration measured with MEG. *Neuroscience Letters, 484*(1), 76-80.

[98]Yang, W., Yang, X., Guo, A., Li, S., Li, Z., Lin, J., . . . Zhang, Z. (2022). Audiovisual integration of the dynamic hand-held tool at different stimulus intensities in aging. *Frontiers in Human Neuroscience, 16*, 968987.

[99]Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247-279.

[100]Liu, B., Jin, Z., Wang, Z., & Gong, C. (2011). The influence of temporal asynchrony on multisensory integration in the processing of asynchronous audio-visual stimuli of real-world events: an event-related potential study. *Neuroscience, 176*(1), 254-264.

[101]Cabeza, R., Anderson, N. D., Locantore, J. K., & Mcintosh, A. R. (2002). Aging gracefully: compensatory brain activity in high-performing older adults. *Neuroimage, 17*(3), 1394-1402.

[102]Greenwood, & M., P. (2007). Functional plasticity in cognitive aging: Review and hypothesis. *Neuropsychology, 21*(6), 657.

[103]Esiri, M. M. (2010). Ageing and the brain. *Journal of Pathology, 211*(2), 181-187.

[104]Ren, Y., Ren, Y., Yang, W., Tang, X., Wu, F., Wu, Q., . . . Wu, J. (2018). Comparison for younger and older adults: Stimulus temporal asynchrony modulates audiovisual integration. *International Journal of Psychophysiology, 124*, 1-11.

[105]Wu, J., Yang, J., Yu, Y., Li, Q., Nakamura, N., Shen, Y., . . . Abe, K. (2012). Delayed Audiovisual Integration of Patients with Mild Cognitive Impairment and Alzheimer's Disease Compared with Normal Aged Controls. *Journal of Alzheimers Disease Jad, 32*(2), 317-328.

[106]Craik, F. I. M., & Salthouse, T. A. (2007). *The Handbook of Aging and Cognition*.

[107]Kolev, V., Falkenstein, M., & Yordanova, J. (2006). Motor-response generation as a source of aging-related behavioural slowing in choice-reaction tasks. *Neurobiology of aging, 27*(11), 1719-1730.

[108]Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Curr Biol, 16*(14), 1422-1427. doi:10.1016/j.cub.2006.05.048

[109]Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Front Psychol, 2*, 264. doi:10.3389/fpsyg.2011.00264

[110]Baddeley, A. D., Hitch, G. J., & Bower, G. H. (1974). The Psychology of Learning and Motivation.

[111]Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience, 9*, 197.

[112]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. doi:10.1037//0278-7393.6.2.174.

[113]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.

[114]Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences, 12*(11), 411-417.

[115]Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping, 25*(1), 46-59.

[116]Ren, Y., Xu, Z., Lu, S., Wang, T., & Yang, W. (2020). Stimulus specific to age-related audio-visual integration in discrimination tasks. *i-Perception, 11*(6), 2041669520978419.

[117]Regenbogen, C., Seubert, J., Johansson, E., Finkelmeyer, A., Andersson, P., & Lundström, J. N. (2018). The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Human brain mapping, 39*(3), 1313-1326.

[118]Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—implications for training and transfer. *Intelligence, 38*(6), 625-635. doi:10.1016/j.intell.2010.09.001

[119]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. doi:10.1007/s00221-004-1913-2.

[120]Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied cognitive psychology, 25*(6), 943-951.

[121]Pahor, A., Collins, C., Smith-Peirce, R. N., Moon, A., Stavropoulos, T., Silva, I., . . . Seitz, A. R. (2021). Multisensory Facilitation of Working Memory Training. *Journal of Cognitive Enhancement, 5*(3), 386-395. doi:10.1007/s41465-020-00196-y.

[122]Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature reviews neuroscience, 17*(7), 438-449.

[123]Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology, 45*(1), 152-170. doi:10.1111/j.1469-8986.2007.00602.x.

[124]Cohen, J. R., Sreenivasan, K. K., & D'Esposito, M. (2014). Correspondence between stimulus encoding-and maintenance-related neural processes underlies successful working memory. *Cerebral cortex, 24*(3), 593-599.

[125]Kitada, R., Sasaki, A. T., Okamoto, Y., Kochiyama, T., & Sadato, N. (2014). Role of the precuneus in the detection of incongruency between tactile and visual texture information: a functional MRI study. *Neuropsychologia, 64*, 252-262. doi:10.1016/j.neuropsychologia.2014.09.028.

[126]Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology, 118*(10), 2128-2148.

[127]Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. *Handbook of aging and cognition*, 155-219.

[128]Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic bulletin & review, 22*(2), 366-377. doi:10.3758/s13423-014-0699-x.

[129]Bogg, T., & Lasecki, L. (2015). Reliable gains? Evidence for substantially underpowered designs in studies of working memory training transfer to fluid intelligence. *Frontiers in psychology, 5*(1589), 1-8. doi:10.3389/fpsyg.2014.01589.

[130]Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Saults, J. S., & Christ, S. E. (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience, 23*(10), 2852-2863.

[131]von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research, 78*(6), 803-820. doi:10.1007/s00426-013-0524-6.

[132]Baum, S. H., & Stevenson, R. A. (2017). Shifts in audiovisual processing in healthy aging. *Current Behavioral Neuroscience Reports, 4*, 198-208.

[133]Shalchy, M. A., Pergher, V., Pahor, A., Van Hulle, M. M., & Seitz, A. R. (2020). N-Back Related ERPs Depend on Stimulus Type, Task Structure, Pre-processing, and Lab Factors. *Front Hum Neurosci, 14*(549966), 1-13. doi:10.3389/fnhum.2020.549966

[134]Peng, Z., Dai, C., Ba, Y., Zhang, L., Shao, Y., & Tian, J. (2020). Effect of sleep deprivation on the working memory-related N2-P3 components of the event-related potential waveform. *Frontiers in neuroscience, 14*(469), 1-9. doi:10.3389/fnins.2020.00469.

[135]Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28*(2), 240-244. doi:org/10.1111/j.1469-8986.1991.tb00417.x

[136]Lawlor-Savage, L., & Goghari, V. M. (2016). Dual N-back working memory training in healthy adults: a randomized comparison to processing speed training. *PloS one, 11*(4), 1-17. doi:10.1371/journal.pone.0151817.

[137]Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of

working memory. *Neuroscience & Biobehavioral Reviews, 26*(2), 105-125. doi:10.1016/s0149-7634(01)00063-x.

[138]Jensen, O., & Lisman, J. E. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends in neurosciences, 28*(2), 67-72. doi:10.1016/j.tins.2004.12.001.

[139]Moran, Z. D., Bachman, P., Pham, P., Cho, S. H., Cannon, T. D., & Shams, L. (2013). Multisensory encoding improves auditory recognition. *Multisensory research, 26*(6), 581-592. doi:10.1163/22134808-00002436.

[140]Oelhafen, S., Nikolaidis, A., Padovani, T., Blaser, D., Koenig, T., & Perrig, W. J. (2013). Increased parietal activity after training of interference control. *Neuropsychologia, 51*(13), 2781-2790. doi:10.1016/j.neuropsychologia.2013.08.012

[141]Schleepen, T. M., & Jonkman, L. M. (2009). The development of non-spatial working memory capacity during childhood and adolescence and the role of interference control: an N-Back task study. *Developmental neuropsychology, 35*(1), 37-56. doi:10.1080/87565640903325733.

[142]Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive psychology, 33*(1), 5-42.

[143]Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. *Principles of frontal lobe function*, 466-503.

[144]Jaušovec, N., & Jaušovec, K. (2012). Working memory training: improving intelligence–changing brain activity. *Brain and cognition, 79*(2), 96-106.

[145]Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., . . . Lisman, J. E. (2001). Gating of human theta oscillations by a working memory task. *Journal of Neuroscience, 21*(9), 3175-3183.

[146]Pergher, V., Wittevrongel, B., Tournoy, J., Schoenmakers, B., & Van Hulle, M. M. (2019). Mental workload of young and older adults gauged with ERPs and spectral power during N-Back task performance. *Biological psychology, 146*, 107726.

[147]Costers, L., Van Schependom, J., Laton, J., Baijot, J., Sjøgård, M., Wens, V., . . . D'hooghe, M. B. (2020). Spatiotemporal and spectral dynamics of multi-item working memory as revealed by the n-back task using MEG. *Human brain mapping, 41*(9), 2431-2446.