FACULTY OF PSYCHOLOGY AND EDUCATIONAL SCIENCES

MULTISENSORY WORKING MEMORY: INFLUENCE OF AMBIGUOUS SOUNDS ON MASKED VISUAL RECOGNITION TASK

Word count: 10175

Laura De Laere
Student number: 01600650

Supervisor: Prof. Dr. Durk Talsma

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Theoretical and Experimental Psychology

Academic year: 2022 – 2023

GHENT UNIVERSITY

Abstract

When performing working memory tasks, multisensory integration reveals multiple benefits. In particular, throughout visual search or recognition tasks, increasing evidence reveals the presentation of semantically congruent auditory stimuli to result in enhanced performances. Semantically incongruent auditory stimuli are revealed to interfere with performances. During the present study, we investigated the presence of the semantic congruency effect when performing a masked visual recognition task combined with ambiguous sounds. Blurred visual stimuli slowly came into focus, and participants had to correctly identify the objects as quickly as possible. Four possible sound conditions were presented at picture onset: sounds congruent with the picture, sounds incongruent with the picture, white noise and no sound. The congruent and incongruent sounds were considered ambiguous, as they lacked easily interpretable characteristics. Overall, results revealed faster reaction times for the congruent sound condition in comparison to all other sound conditions. This resembles the semantic congruency effect. However, the accuracies revealed no differences. Caution is necessary, as a substantial perceptual learning effect was present due to stimulus repetition. When focusing on first presentation only, the semantic congruency effect for reaction times disappeared and only a non-significant trend could be found for the accuracies. As the perceptual learning effect and possible different levels of ambiguity were present, adjustments to the paradigm are necessary. Removing repetition and mapping the different levels of ambiguity, could lead to interesting approaches for the future.

Keywords: multisensory integration, working memory, semantic congruency

Abstract (Dutch version)

Wanneer men werkgeheugentaken uitvoert, kan multisensoriële integratie leiden to verscheidene voordelen. In het bijzonder, tijdens visuele zoek- of herkenningstaken, toont onderzoek aan dat de presentatie van semantisch gelijkaardige geluiden leidt tot betere prestaties. Semantisch ongelijkaardige geluiden zouden interferen met de prestaties. Tijdens deze studie, onderzochten wij de aanwezigheid van het semantische congruentie effect wanneer men een gemaskeerde visuele herkenningstaak uitvoert in combinatie met ambigue geluiden. Vervaagde visuele plaatjes kwamen langzaamaan in focus, terwijl participanten zo correct en snel mogelijk de objecten moesten identificeren. Vier verschillende geluidscondities werden gepresenteerd tijdens de plaatjes: geluiden congruent met het plaatje, geluiden incongruent met het plaatje, white noise en geen geluid. De congruente en incongruente geluiden werden aanschouwt als ambigue, aangezien deze een gebrek toonden aan makkelijk te interpreteren karakteristieken. Over het algemeen, toonden de resultaten snellere reactietijden voor de congruente geluidsconditie in vergelijking met alle andere geluidscondities. Dit lijkt het semantische congruente effect te weerspiegelen. Echter, de correctheid toonde geen verschillen aan. Voorzichtigheid is noodzakelijk, aangezien een substantieel perceptueel leereffect werd gevonden door de herhaling van de plaatjes. Wanneer we enkel focussen op de eerste presentatie van de plaatjes, lijkt het semantische congruente effect voor reactietijden te verdwijnen. Voor de correctheid, kon enkel een niet-significante trend worden gevonden. Door de aanwezigheid van het perceptuele leereffect en de mogelijke verschillende niveaus van ambiguïteit, zijn veranderingen aan het paradigma noodzakelijk. Het verwijderen van de herhaling en het in kaart brengen van de verschillende niveaus van ambiguïteit, kan leiden tot interessante benaderingen voor de toekomst.

Acknowledgement

Writing a master's thesis and performing research, never happens alone. Therefore, it is important to thank all people involved in this fascinating process. First and foremost, I would like to thank prof. dr. Durk Talsma for his time and guidance throughout this project. Whenever I had questions or doubts, he was present to help with encouragement and a smile. His expertise in the field is admirable and infectious. Second, I would like to thank prof. Diane Pecher and student Faye Bettens, for the help and inspiration included in this thesis. And of course, I would like to thank my dear family. My mother & Dirk, my father & Jara, my sister Lisa & Stijn and all of my beloved grandparents. They have shown tremendous love and support throughout this academic journey. Thank you for giving me that extra push whenever it was needed. As have my friends, whom a lot of them even participated in this study. I am forever grateful to have such a wonderful support system.

Index

Abstract **2**

Abstract (Dutch version) **3**

Acknowledgement **4**

Introduction **6**

Multisensory Integration *6*

Multisensory Neurons 6

Neuroimaging 8

Working Memory *9*

Neuroimaging 12

Multisensory Working Memory *13*

Studies on the Multisensory Working Memory 14

Integration of Auditory Information when performing Visual Search or Recognition Tasks 16

The present study *19*

Method **20**

Participants *20*

Stimuli and apparatus *20*

Experimental design and procedure *21*

Data analysis *22*

Results **23**

Additional results *24*

Discussion **26**

Conclusion **30**

References **31**

Appendix **41**

Introduction

Multisensory Integration

In our daily lives, a variety of stimuli is continuously coming our way. We perceive our environment through a variety of senses, including: sight, sound, smell, taste, and touch (Quak et al., 2015). This allows us to properly comprehend our surroundings. Researchers initially assumed that our senses operated largely in isolation and therefor studied them independently (Egeth & Smith, 1967). As many of these different sensory impressions emerge from common occurrences or objects in the external world, combining them appeared be beneficial (Macaluso & Driver, 2005). Around the early 2000s, a renewed interest in the concept of multisensory integration arose. Multisensory integration processes are involved in combining sensory signals, so that they can be processed together in the brain (Stein & Stanford, 2008). At present, sensory information is assumed to interact early on in the sensory processing stream, starting in early cortical areas, and progressing to higher-order processing in the higher-order cortical areas (Recanzone, 2009). This leads to certain modalities influencing one another at various times of processing and resulting in the formation of a unified concept of our environment (Quak et al., 2015). Based on the continuous integration of ongoing sensory input, this mental representation of a unified concept is constantly updated (Talsma, 2015). However, how exactly is multisensory integration implemented in these neural systems?

Multisensory Neurons

A frequently used illustration of how multisensory integration operates, involves multisensory neurons (Meredith, 2002; Talsma et al., 2010). Such neurons are responsive to more than one sensory modality (Talsma et al., 2010). Traditionally, it was believed that multisensory integration only occurred in higher-order cortical areas, such as the frontal cortex, as these association areas are important to form a representation of the outside world (Ghazanfar & Schroeder, 2006). These areas do have connections to multiple early sensory cortical areas, as for example the frontal cortex is connected to both the visual and auditory cortices (Jones & Powell, 1970). Recent studies suggest, however, that some of the neurons in these early sensory cortical areas are also influenced by more than one modality, indicating an early onset of multisensory integration (Recanzone, 2009). These results led to the belief that both the early sensory cortical areas and the high-order cortical areas have a multisensory nature (Ghazanfar & Schroeder, 2006; Recanzone, 2009).

As different sensory receptors can influence one another, their neural connections must come together in shared multisensory neurons (Meredith & Stein, 1986). Thus, for one modality to influence the other, convergence of multisensory information is necessary (Meredith, 2002). Meredith (2002) proposes two types of convergence, excitatory-excitatory or enhancement and excitatory-inhibitory convergence. For the importance of this study, we will only focus on enhancement convergence as this will be relevant for the current study. Enhancement involves multisensory neurons getting stronger activations when multimodal information is provided in comparison to unimodal information (Meredith & Stein, 1986). The influential study by Meredith & Stein (1986) recorded multisensory neurons in the superior colliculus from cats, as they tend to be heavily present in this area. During the presentation of visual, auditory or audio-visual stimuli, neuronal activity was measured. The results revealed that multisensory neurons showed stronger activity to audio-visual stimuli in comparison to separated visual or auditory stimuli. In behavioral studies, multisensory enhancement typically results in enhanced performance revealing faster reaction times, higher accuracies or faster stimulus detection (Laurienti et al., 2004; Meredith, 2002). For example the straightforward study performed by Diederich & Colonius (2004), where participants had to press a button as soon as stimuli appeared. The reaction times decreased when multisensory stimuli were presented compared to the unisensory stimuli. The enhancement is also highly dependent on various factors such as the temporal and spatial relationship of the multimodal stimuli (Meredith, 2002; Talsma et al., 2010). When multisensory stimuli are in temporal and/or spatial alignment, the enhancement enlarges. However, when misalignment is observed, multisensory depression may also occur resulting in weaker responses (Talsma et al., 2010).

Besides these factors, also semantic congruency appears to play an important role during multisensory enhancement. When audiovisual stimuli are semantically congruent (or informationally congruent), the multisensory enhancement in performance enlarges (Tsilionis & Vatakis, 2016). An example of a study, was performed by Suied et al. (2009). Participants simply had to respond during a go/no-go task when target stimuli appeared. These could be visual, auditory or audio-visual. The audio-visual stimulus pairings could either be semantically congruent, the picture of a telephone and the sound of a telephone ringing, or semantically incongruent, the picture of a telephone and the sound of a frog. When the target was for example a frog, participants also had to respond if either the visual or auditory stimulus during multisensory presentation was related to the target. Results revealed faster reaction times for semantic congruent stimuli in comparison to the unisensory stimuli and

semantic incongruent stimuli. This semantic congruency effect, especially for the reaction times, has been replicated multiple times (Iordanescu et al., 2008; Laurienti et al., 2003; Molholm et al., 2004). However, for the accuracies no significant differences were observed in this study, contrary to other studies which will be explained later on (Molholm et al., 2004). This influence of semantic congruency may be important during this paradigm, as the audio-visual stimulus-pairings that will be used can be semantically congruent or incongruent. This could lead to possible differences in reaction times and/or accuracies.

Neuroimaging

The initial multisensory integration research started very early on, around the 1900s (James & Stevenson, 2012). However, the research really took off around the late 1980s and early 1990s, where higher-order cortical areas were the first believed to be multisensory (Talsma, 2015). This was due to the presence of multisensory neurons, their connections to multiple early cortical areas and lesion studies (Ghazanfar & Schroeder, 2006). Three classical multisensory higher-order areas have been proposed; the superior temporal sulcus, the intraparietal complex and the frontal cortex (Cappe et al., 2012; Ghazanfar & Schroeder, 2006). Multiple studies on monkeys provided evidence for multisensory integration in higher-order cortical areas (Andersen et al., 1997; Fogassi et al., 1996; Fuster et al., 2000). An interesting study trying to compare the multisensory brain regions found in monkeys with equivalent areas in human beings, was performed by Bremmer et al. (2001). When conducting their fMRI-study they presented stimuli from different modalities. In three areas - the intraparietal sulcus, the lateral inferior postcentral cortex, and the ventral premotor - they observed enhanced neuronal activity indicative of multisensory integration and multisensory neurons (Meredith, 2002). These areas were equivalent to the areas that were previously found in monkeys. An extensive amount of research considering how multisensory integration operates in higher-order cortical areas testing human beings is now also being provided (Cappe et al., 2012; Regenbogen et al., 2018), as the previous focus was usually on monkeys (Ghazanfar & Schroeder, 2006).

Later on, in the late 1990s in combination with the rise of neuroimaging techniques, a shift in the multisensory integration research appeared (Talsma, 2015). Findings revealing an early temporal initiation of multisensory integration, were the main source of the idea that multisensory integration also occurs during the earlier stages of processing. An example is the ERP-study performed by Giard & Peronnet (1999), which revealed early interactions for the auditory and the visual modality. During this study, participants had to indicate which of two objects was being displayed by pressing a key. The objects either had auditory features, visual

features or audio-visual features. Increasing multimodal-interaction activities were found in the occipitoparietal cortex as early as 40ms after the presentation of the audio-visual stimulus. This ERP-study was one of the first studies that might indicate multisensory integration to occur early on in the sensory processing chain, in brain regions previously considered to be unimodal. Foxe et al. (2000) conducted a similar ERP-study investigating auditory stimuli, somatosensory stimuli and the combination of both. Their results revealed multisensory interactions with an onset of 50ms post-stimulus presentation. These interactions were found over the central and postcentral scalp, which combined with the timing, once more suggests for multisensory integration occurring in brain regions that were previously assumed to be unimodal.

In addition, also functional imaging was used to demonstrate the early emergence of multisensory integration (Driver & Noesselt, 2008). An example is the rather straightforward fMRI-study conducted by Martuzzi et al. (2007), where participants were only required to respond to stimuli by pressing a key. The stimuli could either be visual, auditory or audio-visual. Primarily, multisensory interactions were found in both the visual and the auditory cortices as shifts in BOLD response latencies. Second, the simple presentation of a visual stimulus lead to BOLD amplitude responses in the auditory cortex and vice versa for auditory stimuli in the visual cortex. Considering the evidence of multisensory interactions within early processing stages as previously mentioned (Foxe et al., 2000; Giard & Peronnet, 1999), the results can be interpreted as multisensory interactions instead of mediation (Martuzzi et al., 2007). Studies as such, can also reveal the presence of multisensory neurons (Meredith, 2002). With time, more literature on multisensory integration during early cortical processing emerges and is now a widely accepted concept (Foxe & Schroeder, 2005; Schroeder & Foxe, 2005). As we do accept the idea that the brain operates multisensory in both the higher-order and early cortical areas, it is highly plausible to expect multisensory integration processes in the working memory as well.

Working Memory

When using sensory information coming from our surroundings in a host of cognitive tasks, we typically employ our working memory. Somewhat simplified, our working memory is a system that allows us to temporarily store and manipulate information (Baddeley, 1998). This storage and manipulation of information is necessary to perform tasks such as reasoning, learning and comprehension (Baddeley, 2010). The working memory is characterized by a limited capacity, as many studies refer to the capacity of storing four items at once usually during short-term memory tasks (Cowan, 2001). However, the exact capacity of working

memory is hard to define as different tasks can require different manipulations (Linden, 2007). But how exactly does the working memory operate and how is it implemented in the brain?

One of the first detailed descriptions of how our working memory operates, was provided by Atkinson & Shiffrin (1968). They viewed the human working memory as a single system divided in three different components; the sensory register, the short-term store and the long-term store. Incoming information would be registered in the sensory register according to its sensory modality. Selected information could then be transferred to the short-term store, which also has access to the long-term store. The short-term store would control the flow of information coming in and out of the long-term store. Atkinson & Shiffrin (1968) would describe this short-term store as the 'working memory' where information is also classified according to its sensory modality. These classifications of sensory modalities persisted, while the information was being transferred to the long-term store. However, this model was unable to fully explain the findings during studies with lesion patients (Baddeley, 2010). The famous story of lesion patient H.M. (Corkin, 2002) revealed him not being able to form new ongoing memories as he had an impaired long-term memory. However, he could perform normally on short-term memory tasks such as repeating numbers revealing a preserved short-term memory. In the Atkinson & Shiffrin (1968) model, this dissociation could not be possible as the preserved short-term memory should control the long-term memory. Thus, new adjustments to the model were necessary (Baddeley, 2010).

In 1974, Baddeley & Hitch (1974) provided a new multicomponent model as they wanted to replace the concept of a single system adding temporary sensory buffers. As described in the paper written by Baddeley (2010), the model starts with the central executive, better described as a control system that operates through attention. The central executive was provided with information depending on two short-term memory storages; the visuo-spatial sketchpad and the phonological loop. These two storages store and manipulate, respectively, visual material and the acoustic / verbal information. The dissociations of these two storages were again based on lesion studies, as patients could reveal impairment in one but not in the other (Hanley et al., 1991; Vallar & Baddeley, 1984). Nowadays, the multicomponent model has added an extra storage system to help the central executive; the episodic buffer (Baddeley, 2010). When the multisensory integration account arose, it was proposed that the central executive played an important role but lacked a short-term multimodal store (Baddeley, 2000). The added episodic buffer would store the multisensory information, rather than

sensory information segregated according to its modality, and it would also interact with long-term memory information (Quak et al., 2015).

Although the multicomponent model from Baddeley & Hitch (1974) remains influential, however, a variety of models have since been proposed and gaining prominence (D'Esposito & Postle, 2015). D'Esposito & Postle (2015) identified these new models as state-based models, where the allocation of attention appears to be of importance. D'Esposito & Postle (2015) divided these models in two categories. The first category involves models where semantic representations in long-term memory are activated. An important model in this category was proposed by Cowan (1995). As explained by Cowan (1999), this model consisted of three components; the long-term store, the short-term store and the focus of attention. In this model, the portion of the long-term store that had a higher level of activation was called the short-term store. Due to limited capacity (Cowan, 2001), part of the short-term store could then be in the focus of attention. For this specific model, the central storage would be multimodal (Quak et al., 2015). Already implying Postle's (2006) proposal, Baddeley (2000) found some similarities between the multicomponent model and the embedded model proposed by Cowan (1995). It could be interpreted as an interface between the central executive, limited by attention, and the storage-limited episodic buffer (Baddeley, 2000).

The second category of models are those involving systems that perceive information can also contribute to the short-term retention of that information. This is usually based on and referred to as the sensorimotor recruitment model or theory (Scimeca et al., 2018). For example, when performing visual working memory tasks, research suggests the early visual areas to play a role in supporting the visual working memory representations (Rademaker et al., 2019; Yörük et al., 2020). An example study of this theory, will be explained later on. This account is still supported, but now in a more flexible manner as other areas may also support the working memory representations (Yörük et al., 2020). As there are numerous models in both categories and literature on working memory is extensive, we like to refer to the review paper by D'Esposito & Postle (2015) for more information.

Nowadays, there is an ongoing discussion on whether the latest models, such as the multicomponent model, are still adequate to explain the increasing amount of new empirical data. Postle (2006) argues that a new model should be created that incorporates brain regions with sensory, action-related, and representational capabilities that could operate through attention. Simplified, it could resemble an integration of all previously mentioned models (D'Esposito & Postle, 2015) as each model has its own pros and cons. The part of Postle's (2006) proposition dealing with sensory perception is the most intriguing for the purposes of

the current study. As the current empirical data suggests that brain systems having a sensory function can also be involved for the short-term storage of this exact information, which will be discussed in even more detail later on. This also leads to the great probability of a multisensory working memory (Quak et al., 2015).

Neuroimaging

Using neuroimaging techniques such as fMRI and TMS combined with electrophysiology, studies are trying to find brain areas that can implement working memory processes (Chai et al., 2018). When performing working memory tasks the fronto-parietal brain regions become activated (Linden, 2007). These regions include the dorsolateral prefrontal cortex, the anterior cingulate cortex and parietal cortices. Working memory does probably involve the functional integration of the brain as a whole (Chai et al., 2018), however, it has been established that the posterior parietal cortex executes the sensory and perceptual processing (Andersen & Cui, 2009). This may also involve multisensory processing (Andersen et al., 1997), insinuating at a multisensory working memory, as will be discussed later on. The dorsolateral prefrontal cortex would function as the central executive and the anterior cingulate cortex as the attention controller (Chai et al., 2018). An example study was conducted by Oliveri et al. (2001). Using TMS, the authors tried to interfere with the functioning of parietal, temporal and frontal regions during a visual working memory task. When they interfered with the parietal cortex (bilateral), it disrupted the performance for visual-spatial working memory tasks with increased reaction times. When interfering with the temporal cortex, it disrupted the performance on visual-object tasks with increased reaction times. These results do suggest the involvement of the posterior cortex in perceptual processing, for visual spatial and visual object information. When they interfered with the dorsolateral prefrontal cortex, it disrupted the performance during both the visual-spatial and visual-object tasks revealing increased reaction times and lower accuracies. Such results do suggest the frontal region to function as a multisensory monitoring area which could be interpreted as the central executive. In addition, subcortical regions are also being implicated as involved in the working memory (Chai et al., 2018). Consider, for example the study by Moore et al. (2013), involving the basal ganglia. Given this trend, it is reasonable to suppose that not all of the regions involved in working memory have been discovered yet.

As was discussed before, empirical evidence suggests brain regions which are necessary for sensory processing to also provide the short-term storage of such information (Postle, 2006; Quak et al., 2015). During an example study by Yörük et al. (2020), participants had to remember the orientation and the precise location of visual working

memory items consisting of three colored bars. There were two sorts of trials, orientation-report and location-report trials where participants, respectively, had to adjust the orientation or location of three bars to match the target display as presented before. It was found that during both working memory tasks, typical results of reliance on the early visual cortex were present. Other studies found similar results, as for example the fMRI-study by Zhao et al. (2022). Here, participants performed only delayed orientation recall tasks. Visual representations were found in the contralateral and ipsilateral primary visual cortex or V1. These findings point to the previously described sensory recruitment theory (D'Esposito & Postle, 2015). However, nowadays, it is suggested to follow a more flexible form of this theory as other regions can store this information as well (Xu, 2018; Yörük et al., 2020).

Multisensory Working Memory

Now the question still arises if there is a connection between multisensory integration and working memory. This paper follows the multisensory integration account which is based on the concept of multisensory integration occurring at various stages during the processing of sensory information (Koelewijn et al., 2010; Macaluso & Driver, 2005; Talsma et al., 2010). A relationship between this multisensory integration account and working memory must be established to conduct research based on the premise of a multisensory working memory.

Many cognitive tasks involve the temporary storage of sensory information (Pasternak & Greenlee, 2005). As already discussed, Postle (2006) proposes that when the brain can represent a certain type of sensory information, this information is also temporarily be retained in that area (Zhao et al., 2022). Features of stimuli are now found to be maintained in feature-selective systems that not only include the prefrontal and parietal cortex but also the sensory areas where the early processing takes place (Pasternak & Greenlee, 2005). As was discussed before, it has now been established that multisensory integration also occurs during early stages of processing in the early sensory cortices (Recanzone, 2009). Taking these two occurrences into consideration, Quak et al. (2015) proposed that there might also be a possibility that multisensory information is stored as a unified representation in our working memory. Moreover, multisensory integration and working memory processes also appear to occur in similar brain areas. Multisensory integration was revealed during monkey studies in both the prefrontal and the parietal cortex (Andersen et al., 1997; Romanski & Hwang, 2012), which are both areas part of the large fronto-parietal working memory network (Linden, 2007). A multisensory working memory is thus highly plausible, however, scientific evidence for such a working memory system is currently still relatively sparse.

Studies on the Multisensory Working Memory

If a multisensory working memory is a plausible concept, scientific evidence must point in this direction. Based on this rather novel concept of a multimodal working memory, additional studies have been conducted as this provided a new approach in research. An early study testing multisensory working memory was conducted by Thompson & Paivio (1994). Participants saw either visual, auditory or audio-visual stimulus-pairings and had to perform a free recall test afterwards. The results revealed that participants were significantly better at remembering audio-visual stimuli in comparison to either the visual or auditory stimuli. Similar results have been found during many other studies as well. Another example is the study performed by Goolkasian & Foos (2005) where participants had to recall 3 or 6 items while also evaluating the accuracy of mathematic sentences. When the to-be-remembered pictures were accompanied by matching spoken words, enhanced performances were revealed. Interestingly, when the spoken words were incongruent, this did interfere with participants' performance. It is rather clear that multisensory information in general usually leads to improved performances throughout working memory paradigms compared to unisensory information (Meredith, 2002; Quak et al., 2015).

One of the most recent papers supporting the notion of a multisensory working memory was produced by Pahor et al. (2021). They hypothesized that multisensory WM training would transfer better to new, untrained WM tasks then just a visual training or a visual and an auditory training that was presented separately. Participants in this study were assigned to one of four conditions: visual WM training, alternating visual and auditory WM training, multisensory WM training and a control condition. The multisensory stimuli used in the third condition always consisted of congruent pairs of stimuli. After completing their training, participants had to perform untrained WM tasks to determine which training transferred best. The results demonstrated that the visual training and the multisensory training resulted in the greatest gain of performance on the tasks. However, when it came to some of the untrained tasks, the multisensory training showed significant transference. The three training groups outperformed the control group in general, but when it came to the complex span tasks the multisensory training group once again demonstrated significant performance improvement. In conclusion this study proves that multisensory integration training might improve working memory during certain tasks, highlighting the benefits of a possible multisensory working memory (Pahor et al., 2021; Shams & Seitz, 2008).

For specific aspects of working memory, such as the visuo-spatial working memory (VSWM), multisensory information appears to benefit performance as well (Botta et al.,

2011). In the Botta et al. (2011) study, participants had to recall arrays consisting of colored squares. After the memory array presentation, a test array was presented with one marked square. Participants then had to respond if the indicated square matched the color as was displayed in the memory array. Prior to the display of any array, a spatial cue - which could be either visual, auditory or both - was given to indicate the spatial location of the marked square. When the multisensory cues were spatially congruent, they revealed higher attentional effects compared to the spatially congruent unisensory cues. The authors say that multisensory information may increase the biasing of information access in VSWM through spatial attention.

In addition to the behavioral studies, neuroimaging studies on multisensory working memory have been conducted as well. Based on previous studies, the intraparietal sulcus functions as a region for visual working memory maintenance (Todd & Marois, 2004; Xu & Chun, 2006). Considering Cowan et al.'s (2011) work, however, this may also be a region where multisensory integration and working memory functions are located. Cowan et al.'s (2011) fMRI study consisted of two experiments. The first involved a memory task, where either visual or audio-visual stimuli had to be remembered and compared to a probe item. The visual condition existed of a smaller and a heavier load. The results revealed nine potential regions of interest where the activity increases during the heavier loaded visual condition and the multimodal audio-visual condition. This could lead to brain regions responding to both visual and auditory memory loads. During the second experiment, they added a small-load auditory condition. Now all of the stimuli had to be remembered, as no probe was needed and a suppression task was added. The results revealed an increased activity of the left intraparietal sulcus whenever memory load was increased, regardless of the modalities of the stimuli. This area was found for both experiment 1 and 2. The left intraparietal sulcus may be perceived as a region for working memory storage for items in multiple modalities, as was also proposed by others such as Majerus et al. (2010).

It is evident from investigating the current research, that multimodal information has certain advantages in comparison to unisensory information during working memory tasks. Recalling multisensory information appears to result in enhanced performances (Goolkasian & Foos, 2005; Thompson & Paivio, 1994), as does the performance on working memory tests after multisensory working memory trainings (Pahor et al., 2021). More specific forms of working memory, such as the visuo-spatial working memory, also appear to benefit from multisensory information (Botta et al., 2011). Studies like these show multisensory interactions which are difficult to explain with only the classic modal models (Quak et al.,

2015). Besides this, multisensory and working memory processes do appear to share brain areas such as the prefrontal and parietal cortex (Andersen et al., 1997; Linden, 2007; Romanski & Hwang, 2012) and in particular the intraparietal sulcus may function as a working memory storage for multisensory information (Cowan et al., 2011; Majerus et al., 2010). Taken together, this study does follow the belief that the working memory operates multisensory.

Integration of Auditory Information when performing Visual Search or Recognition Tasks

There has been a lot of research specifically investigating the integration of auditory and visual modalities (Thelen et al., 2012; Welch & Warren, 1981). Here we specifically focus on visual recognition or visual search working memory tasks mixed with sound conditions. An interesting study investigating semantic congruency and multisensory benefits during a working memory task, was performed by Molholm et al. (2004). Participants had to perform an animal recognition task, where during several blocks the target stimulus changed. For example, in one block the target stimulus was a dog. When the presented stimulus was the target, participants simply had to press a button. The presented stimulus could be visual, auditory or audio-visual. For the audio-visual stimuli, the pairings could either be semantically congruent, the picture of a dog and the sound of a bark, or semantically incongruent, the picture of a dog and the sound of a cow 'lowing'. Overall, when multisensory stimuli were presented participants responded faster. An enhanced performance was also found when the multisensory stimuli were semantically congruent resulting in faster reaction times and higher accuracy rates. However, the semantically incongruent trials did not significantly interfere with participants' performance. Similar results have been replicated repeatedly, specifically for reaction times, leading to the assumption that semantically congruent audio-visual stimuli appear to improved performance. (Laurienti et al., 2004; Suied et al., 2009; Tsilionis & Vatakis, 2016).

Another important study in this field was conducted by Iordanescu et al. (2008). This study aimed to investigate the identification of objects through audio-visual integration. These authors hypothesized that hearing the characteristic sound of a certain object would facilitate the visual search of that particular object. The procedure consisted of a visual search display, where four common objects were presented. To avoid response bias, participants were instructed on which item to search for before to the start of each trial. Three sound conditions that were used were paired with the visual search task; sounds that were consistent with the target object, sounds that were consistent with a distractor object or sounds that were unrelated to all of the displayed objects. The results confirmed the original hypothesis, target-

consistent sounds facilitated the visual search for the target object revealing faster reaction times. Further, the reaction times indicated no differences between the distractor-consistent sounds and the unrelated sounds, which could be explained by the lack of goal-directed top-down feedback. However, for the accuracies no significant differences were found. To account for semantic activation and the distracting properties of the distractor-consistent and unrelated sound condition, the researchers conducted two control experiments. For the first control experiment, the results showed no benefit for characteristic sounds when searching for an object name instead of pictures. For the second control experiment, once more the results provided evidence that target-related sounds facilitated the visual search for the reaction times in comparison with both the distractor-related condition and the added no sound condition. The distractor-related sound condition had no difference in reaction time compared to the no sound condition, ruling out the distracting properties as the explanation of the findings. As during the first experiment, no significant differences for the accuracies were found. This is in line with the results from a previously mentioned study by Suied et al. (2009), however, some studies do find higher accuracies as well (Molholm et al., 2004). In conclusion, this paper provides evidence that audio-visual integration is present when performing a visual search working memory task. It also highlights the importance of semantic congruency of the audio-visual stimuli when focusing on the reaction times.

One of the sources of inspiration for this research idea, was the study performed by Maezawa et al. (2022). They used a visual search task inspired by Iordanescu et al. (2008). Instead of using objects, the authors used pictures of certain events or actions which could be described as verbs, as for example sneezing. Participants had to indicate the location of a target as quick as possible while four events where being presented, associated with certain sounds conditions. The possible sound conditions consisted of a condition congruent with the target, a condition congruent with a distractor and a control condition. The control condition varied amongst the different experiments that they conducted; consisting of sounds unrelated to all the stimuli, white noise and no sound. They hypothesized target-congruent sounds to facilitate the visual target localization and the target-incongruent sounds to impair the visual target localization. Before each experiment was conducted, a familiarity task was presented for the participants to associate the audio-visual pairings. Experiment 1 used the unrelated sound condition, as results revealed the task to be facilitated with faster reactions times (RTs) when being paired with target-congruent sounds in comparison to the other two sound conditions. The RTs of the incongruent and unrelated sound conditions were found to be comparable. Trying to exclude semantic influences, they conducted experiment 2 using the

white noise sound condition. Results revealed that the target-congruent condition resulted in faster RTs compared to the other two sound conditions. These results might reflect auditory enhancement, where complementary auditory stimuli facilitate the visual search task. Furthermore, the RTs for the distractor-congruent sound condition was enlarged compared to the white noise condition. This might reflect semantic influences, as the distractor-congruent condition appears to interfere during the visual search task. Experiment 3 was conducted using the no sound condition to control for the distracting properties of white noise. This experiment replicated both the auditory enhancement and the interference of visual searches, as target-congruent trials revealed again faster reaction times. The no sound condition was responded to faster compared to the distractor-congruent condition. The last three experiments they conducted used the same control conditions, respectively, as to the previously performed ones, but the new experiments never presented identical pictures to control for the ability to learn. As the familiarization task previously did use the same pictures, this was a possibility. However, the results replicated the pattern of those that were previously found. In conclusion all of the results reflected that visual search was facilitated for the RTs when target-congruent sounds were presented compared to conditions where no congruent sounds were present. However, the search performance was impaired when distractor-related sounds were presented compared to white noise and no sound trials. Important to note, is that throughout this study the focus was mainly on the reaction times and not the accuracies. This study implies that multisensory enhancement takes place as audiovisual stimuli help to perform better in a visual search task regarding the reaction times, not only for objects (Iordanescu et al., 2008) but also for event scenes.

Final, the last study we wanted to present has a more similar task paradigm as ours compared to the previously described studies. During the study performed by Chen & Spence (2010), which they based on similar literature (Iordanescu et al., 2008; Molholm et al., 2004), participants had to identify masked pictures. Throughout each trial, pictures were briefly presented (27ms) after which a mask was displayed until the next trial. For the importance of our paradigm, only experiment 1 and 2 will be presented as the others involve the spatial alignment of the audio-visual stimuli. For these experiments, only the accuracies were evaluated. In experiment 1, the pictures could be accompanied by congruent, incongruent or white noise sounds. The results revealed congruent trials to have higher accuracies compared to both incongruent and white noise trials. However, the white noise trials did reveal higher accuracies in comparison to incongruent trials. This resembles the auditory enhancement and the interference effect (Maezawa et al., 2022). For experiment 2, the white noise condition

was replaced with no sound. Once again, the congruent trials revealed higher accuracies compared to incongruent and no sound trials. When comparing the incongruent and no sound trials, no significant differences were found. Taken together, with the literature revealing faster reaction times (Iordanescu et al., 2008; Molholm et al., 2004), it appears that audiovisual semantic congruency influences both the reaction times and the accuracies.

The present study

This present research will not use a visual search paradigm but a visual recognition paradigm. The paradigm is partly based on the one used during an internship performed by Bettens (2022) at the University of Ghent. In this internship, a masked visual paradigm was used where visual stimuli are presented completely degraded at the beginning of each trial, as they gradually sharpen and reveal color until it is fully presented. The trials started with only black and white images excluding any effects due to color. In the internship, the stimulus was accompanied by two pictures on each side to manipulate the effects of context and no sound conditions were present. During the present study, we only used the visual masked paradigm where our single visual stimulus was first completely degraded and slowly coming into focus and revealing its colors. No context pictures were present. Participants will have to recognize and report the visual stimulus as quick as possible by pressing a button. While using this visual masked paradigm, pairings with sound conditions will be made based on the conditions of Maezawa et al. (2022). Four main sound conditions will be used; congruent sounds related to the visual stimulus, incongruent sounds unrelated to the visual stimulus, white noise and no sound condition. In addition, all of the sounds used will be ambiguous sounds which means that the sounds used during this paradigm lack easily interpretable characteristics. This results in ‘masked’ congruent and incongruent sounds conditions, which are used to examine if there is a facilitation effect for the congruent sound condition when the sound is not completely comprehended. We do believe this to be a new approach. This study aims to deliver novel and intriguing findings by switching the design from a visual search task to a visual masked recognition task and employing ambiguous sounds to see if the facilitation of semantically congruent pairings remains present.

Based on this research proposal and previous studies, multiple hypotheses can be formed. The overall hypothesis of this study, is that congruent audio-visual stimuli will result in facilitated visual stimulus recognition with faster reaction times and higher accuracies (Chen & Spence, 2010; Iordanescu et al., 2008; Maezawa et al., 2022; Molholm et al., 2004). This could be possible due to the auditory enhancement effect, where congruent sounds result in enhanced visual performances. Second, also supported by the presented literature, is that

the distracting ambiguous sounds, or the incongruent condition, may impair visual identification resulting in longer reaction times and lower accuracies compared to the white noise condition. At last, the no sound condition is hypothesized to be a neutral condition with no facilitation or impairment of the visual recognition in comparison to the incongruent sound condition (Chen & Spence, 2010; Iordanescu et al., 2008).

Method

Participants

62 participants (mean age = 20.97, range = 17-28; female = 48, male = 14) enrolled in this study via the UGent Sona platform receiving course credits, and also personal contacts which were all highly appreciated for their time and effort to participate. 5 additional participants were tested but discarded due to inefficient performance based on the average response rate (less than 1.5 SD below average). After this, another 5 participants were tested to increase the power during a second testing moment resulting in our sample of 62 participants. All subjects were native Dutch speakers, with normal or corrected vision and no hearing deficiencies. Participants signed an informed consent beforehand and received a debriefing with the results after the study. Using Pangea, we have an estimated power of 92% for the Repeated Measures ANOVA analysis with a Cohen's d effect size of 0.25 with our sample of 62 participants.

Stimuli and apparatus

80 different visual and auditory stimuli were selected from a database that was kindly provided by Prof. dr. Diane Pecher, of the Erasmus school of Social and Behavioural Sciences at the Erasmus University, Rotterdam. 15 participants evaluated the recognizability of the auditory stimuli without the presentation of the corresponding visual stimuli. The 80 stimulus-pairings we chose for this paradigm were recognized in less than 40% of the time. The stimuli were first completely blurred without any color. At the end of each trial, the stimuli were presented in full focus and colored on a white background on the computer screen to increase visibility. For the auditory stimuli wired headphones were provided by us, as devices using Bluetooth can cause a temporal delay which is to be avoided. Stimulus presentation and data collection were performed using PsychoPy (v2022.2.4) on standard PCs in an experimental room with no other distractions present. The distance between the participants and the PC was the standardized 60 cm.

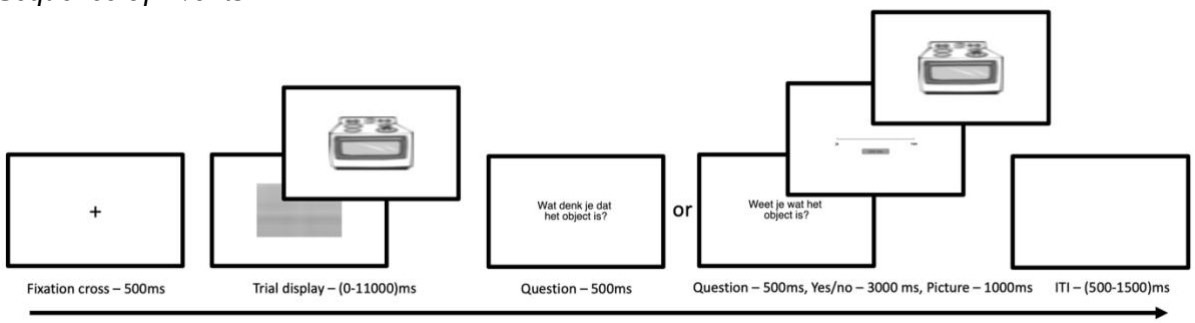
Experimental design and procedure

This study consisted of a masked visual recognition task, where a blurred visual stimulus slowly came into focus and participants had to respond as soon as they recognized the visual stimulus. Each trial was preceded by a fixation cross to indicate the start of a new trial. When participants recognized the stimulus, they had to press the 'h' button which was selected as it is positioned in the middle of the keyboard to avoid handedness biases. As participants pressed the button, the question 'What do you think the object is?' was shortly presented and answers could be typed and locked in by pressing the 'enter' button. If the 'h' button was not pressed during the trial, the question 'Do you know what the object is?' was shortly presented and participants could select yes or no. When 'yes' was selected, participants could once again just type their answer and lock it in by pressing the 'enter' button. When participants selected 'no', the picture was shortly presented in full focus to make sure it was recognizable at all. After this short presentation, participants could type and lock in their answer of what the object was.

Throughout this paradigm, one of four sound conditions could be present during the visual display. The auditory stimulus could be semantically congruent to the visual stimulus, as for example the sound of a hair dryer blowing was accompanied by the picture of a hair dryer. The auditory stimulus could be semantically incongruent to the visual stimulus, for example the sound of a hair dryer blowing accompanied by the picture of a telephone. For the last two sound conditions either white noise or no sound was presented. All sound conditions were counterbalanced, as each of these conditions was displayed in one quarter of all trials.

The sequence of events in a trial was the following (see Figure 1): first a fixation cross appeared and lasted for 500ms. The target display was then presented and could last until 11000ms if no button was pressed. The questions that were presented lasted for 500ms and participants could respond until they pressed the 'enter' button. The 'yes' or 'no' option could be selected during 3000ms. When participants pressed 'no', the picture was shown for 1000ms. After each trial, an inter-trial interval was presented with a random duration from 500 to 1500ms.

Figure 1
Sequence of Events



Note. Example of a trial starting with a fixation cross (500ms) followed by the trial display, with a duration that depends on whether the 'h' button was pressed (0-11000ms). When the 'h' button was pressed, the question 'What do you think the object is?' was briefly presented (500ms). If the 'h' button was not pressed, the question 'Do you know what the object is?' was briefly presented (500ms). Participants could answer with either yes or no (3000ms). When the 'no' option was selected, a short presentation of the visual stimulus was displayed (1000ms). The trial ended with an inter-trial interval (ITI, 500-1500ms).

Overall, the experiment consisted of 160 trials divided into two blocks with 80 trials each to provide participants with a break. As there were 160 trials, all of the visual stimuli were presented twice. During each block, in a quarter or 20 of the 80 trials each previously mentioned sound condition (congruent, incongruent, white noise, no sound) was presented. With the instructions and the small break in the middle, the experiment lasted for approximately one hour.

Data analysis

The behavioral analyses were carried out using JASP (Love et al., 2019) and RStudio (Allaire, n.d.). To control for learning effects (Furmanski & Engel, 2000; Karni & Sagi, 1991), the first and second presentation of the stimuli were added into the analysis in the Repetition variable. The experiment had two main within-subject factors: Sound Condition with 4 levels (Congruent/Incongruent/WhiteNoise/NoSound) and Repetition with 2 levels (First/Second) which were entered into a 4x2 Repeated Measures ANOVA to compare performance across all conditions separately for the reaction times (RTs) and the accuracies. Trials with RTs above or below 2.5 SD from the participant's mean separately for each condition (0.01% overall) were also removed. Accuracies were manually graded based of the response on the presented visual stimulus (0 when incorrect, 1 when correct). Post-hoc paired t-tests were used to evaluate the direction of the main and interaction effects. As for both RTs and accuracies the Repetition variable was highly significant, two additional Repeated Measures ANOVAs were conducted for the variable of Sound Condition separately for the First and Second presentation of the stimuli.

Results

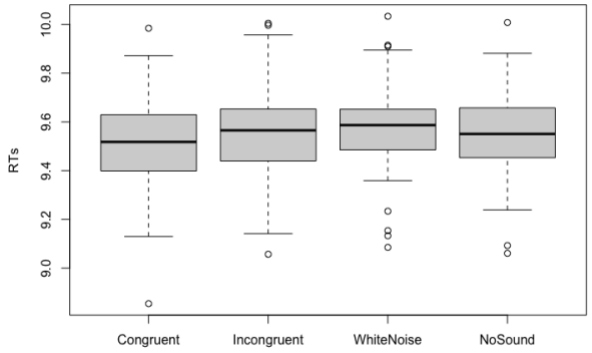
Table 1
Descriptives

		Reaction Times				*Accuracies*			
Sound Condition	**Repetition**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Congruent	First	9.652	0.246	8.539	10.079	0.870	0.083	0.647	1.000
	Second	9.418	0.236	8.818	9.949	0.871	0.091	0.632	1.000
Incongruent	First	9.695	0.236	8.548	10.097	0.836	0.089	0.647	1.000
	Second	9.483	0.245	8.677	10.010	0.872	0.089	0.571	1.000
White Noise	First	9.719	0.197	8.899	10.090	0.842	0.105	0.455	1.000
	Second	9.505	0.202	8.883	9.997	0.869	0.074	0.700	1.000
No Sound	First	9.730	0.142	9.469	10.163	0.824	0.098	0.563	1.000
	Second	9.464	0.217	8.854	9.947	0.905	0.074	0.650	1.000

Note. Divided for the four different levels of Sound Conditions and the two levels of Repetition; the mean, the standard deviation, the minimum and the maximum values are given for both the reaction times and the accuracies.

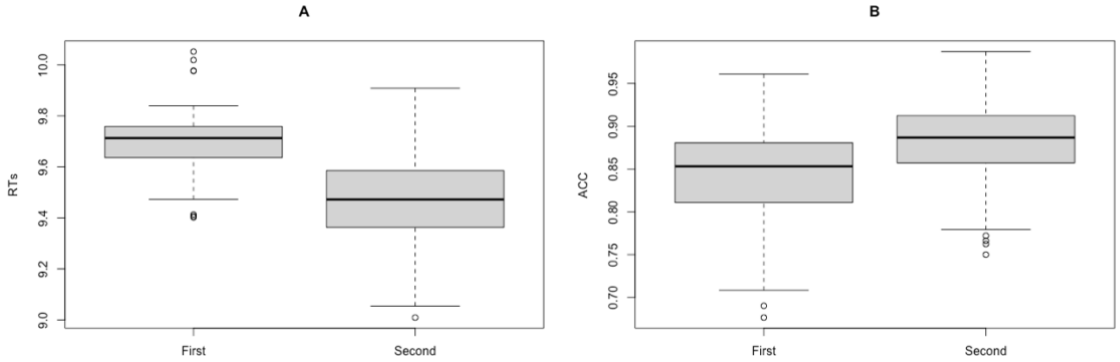
Performing a 4x2 Repeated Measures ANOVA for Sound Condition and Repetition on the reaction times revealed a significant main effect for Sound Condition ($F_{3,183} = 4.868$, $p = 0.003$, $\eta p2 = 0.074$) and a highly significant main effect for Repetition ($F_{1,61} = 126.893$, $p < 0.001$, $\eta p2 = 0.675$) (see Appendix). Post-hoc comparisons focusing on Sound Condition revealed significant differences for Congruent trials in comparison to Incongruent trials ($p = 0.051$), to White Noise trials ($p = 0.003$) and to No Sound trials ($p = 0.022$) (see Appendix). Congruent trials were responded to faster ($M = 9.496$, $SD = 0.196$) in comparison with Incongruent trials ($M = 9.556$, $SD = 0.188$), with White Noise trials ($M = 9.570$, $SD = 0.174$) and with No Sound trials ($M = 9.553$, $SD = 0.170$) (see Table 1). Focusing on the main effect of Repetition, the First presentation trials were responded to slower ($M = 9.701$, $SD = 0.130$) in comparison to the Second presentation trials ($M = 9.466$, $SD = 0.185$) (see Table 1). Results revealed no significant interaction effect for Sound Condition * Repetition ($F<1$). Repeating the same analysis for the accuracies revealed a highly significant main effect for Repetition ($F_{1,61} = 70.598$, $p < 0.001$, $\eta p2 = 0.463$) but no main effect for Sound Condition ($p>0.05$) (see Appendix). The overall accuracy for the First presentation trials was lower ($M = 0.844$, $SD = 0.063$) in comparison to the Second presentation trials ($M = 0.879$, $SD = 0.057$) (see Table 1). Results also revealed a significant interaction effect for Sound Condition * Repetition ($F_{3,183} = 5.779$, $p = 0.002$, $\eta p2 = 0.087$). The post-hoc comparisons revealed several interactions to be significant (see Appendix).

Figure 2
Boxplots for the Main Effect of Sound Condition (RTs)



Note. Boxplots showing the main effect of Sound Condition for the reaction times, where congruent trials were responded to faster in comparison to all other sound conditions. The lines inside the boxes represent the overall medians, the lines limiting the boxes represent the first and third quantiles. The dots represent outliers.

Figures 3 and 4
Boxplots for Repetition



Note. (A) Boxplots showing the main effect of Repetition for the reaction times, where the second presentation revealed faster reaction times. (B) Boxplots revealing the main effect of Repetition for the accuracies, where the second presentation resulted in higher accuracies. The lines inside the boxes represent the overall medians, the lines limiting the boxes represent the first and third quantiles. The dots represent outliers.

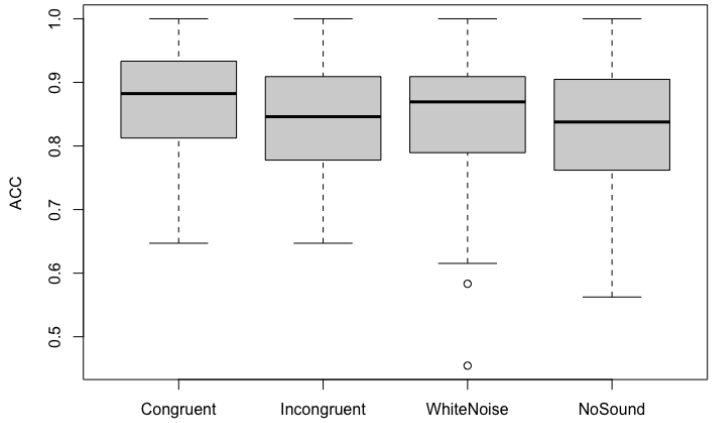
Additional results

Only focusing on the first presentation of the stimuli, a Repeated Measures ANOVA was performed for Sound Condition on the reaction times and revealed no significant main effect ($p > 0.05$). Repeating this analysis for the accuracies did reveal a significant main effect for Sound Condition ($F_{3,183} = 3.876$, $p = 0.010$, $\eta p2 = 0.060$) (see Appendix). Post-hoc comparisons revealed a significant difference for Congruent trials vs No Sound trials ($p =$

0.007), where the Congruent trials showed higher accuracies (M = 0.870, SD = 0.083) compared to the No Sound trials (M = 0.824, SD = 0.098) (see Table 1).

Figures 5

Boxplots for the Main Effect of Sound Condition during the First Presentation (ACC)

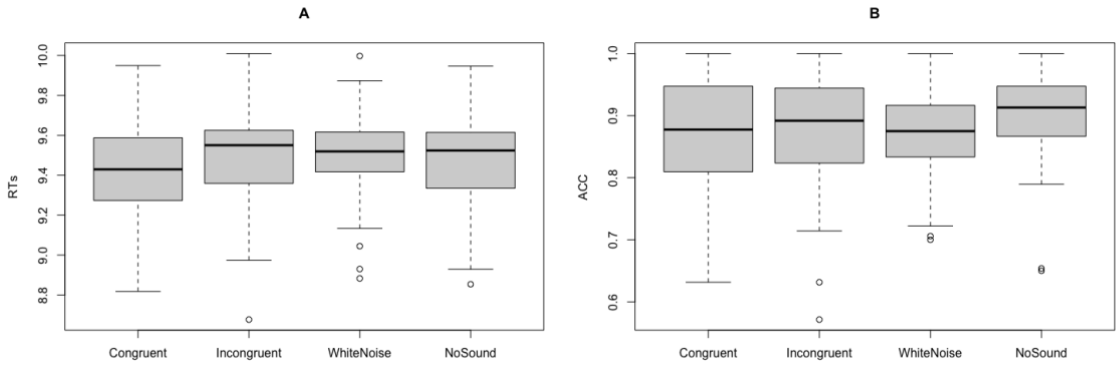


Note. Boxplots showing the main effect of Sound Condition during the first presentation for the accuracies, where congruent trials revealed higher accuracies compared to no sound trials. The lines inside the boxes represent the overall medians, the lines limiting the boxes represent the first and third quantiles. The dots represent outliers.

For the second pair of additional analyses only focusing on the second presentation of the stimuli, or the repetition, a Repeated Measures ANOVA was again performed for Sound Condition on the reaction times. Results revealed a main significant effect for Sound condition ($F_{3,183} = 3.777$, $p = 0.012$, $\eta p2 = 0.058$). Post-hoc comparisons revealed a significant difference for Congruent vs White Noise trials ($p = 0.009$) (see Appendix), where the Congruent trials were responded to faster (M = 9.418, SD = 0.236) in comparison with the White Noise trials (M = 9.505, SD = 0.202) (see Table 1). Repeating this analysis for the accuracies also revealed a significant main effect for Sound Condition ($F_{3,183} = 3.974$, $p = 0.009$, $\eta p2 = 0.061$) (see Appendix). Post-hoc comparisons revealed significant differences for No Sound trials in comparison to Congruent trials ($p = 0.028$), to Incongruent trials ($p = 0.032$) and to White Noise trials ($p = 0.022$) (see Appendix). Overall, the No Sound trials revealed higher accuracies (M = 0.905, SD = 0.074) in comparison with the Congruent trials (M = 0.871, SD = 0.091), with the Incongruent trials (M = 0.872, SD = 0.089) and with the White Noise trials (M = 0.869, SD = 0.074) (see Table 1).

Figures 6 and 7

Boxplots for Main Effects of Sound Condition during the Second Presentation



Note. (A) Boxplots showing the main effect of Sound Condition for the reaction times during the second presentation, where congruent trials were responded to faster in comparison to white noise trials. (B) Boxplots revealing the main effect of Sound Condition for the accuracies during the second presentation, where no sound trials revealed higher accuracies in comparison to all other sound conditions. The lines inside the boxes represent the overall medians, the lines limiting the boxes represent the first and third quantiles. The dots represent outliers.

Discussion

The main goal of this study was to determine whether the well-known semantic congruency effect could also be found using a masked visual recognition task paired with more ambiguous sounds. Three hypotheses were made. First, the congruent sound condition was hypothesized to result in enhanced performances with faster reaction times and higher accuracies compared to all other conditions. Second, the incongruent sound condition was hypothesized to result in impairment of performance with slower reaction times and lower accuracies compared to the white noise condition. Final, the no sound condition was hypothesized to be neutral in comparison to the incongruent condition with similar reaction times and accuracies. For the overall reaction times, this effect appears to be found as congruent trials were responded to faster compared to all other trials. This appears to confirm our first hypothesis, where congruent trials are responded to faster in comparison to the trials of all other conditions, and our third hypothesis. For the overall accuracy, no differences for the sound conditions were found. These findings are not consistent with our first and second hypotheses. It does appear to confirm our third hypothesis as the no sound condition appears to be neutral compared to the incongruent condition. However, we need to remain cautious as for both the reaction times and the accuracies a substantial perceptual learning effect (Furmanski & Engel, 2000; Karni & Sagi, 1991) was found. During the second presentation of the visual stimuli, the reaction times got significantly faster and the accuracies got

significantly higher. Besides this, specifically for the accuracies, the sound conditions resulted significant differences when comparing the first and second presentation of the visual stimuli. For this reason, reaction times and accuracies were analyzed separately for the first and the second presentation (or repetition) of the visual stimuli. Throughout the first presentation of the stimuli, only the accuracies showed significant differences for the sound conditions. The congruent trials revealed higher accuracies compared to the no sound trials. These results partially confirmed our first and third hypothesis. For the second presentation of the stimuli, both the reaction times and the accuracies revealed significant differences for the sound conditions. Focusing on the reaction times during repetition, our first hypothesis was partially confirmed as the congruent trials showed faster reaction times compared to the white noise trials. For the accuracies during repetition, the results were inconsistent with all three of our hypotheses as the no sound trials revealed higher accuracies in comparison to all other sound conditions. As some intriguing results were found during all these analyses, we will discuss these now in more detail below.

When analyzing the overall reaction times, a similar pattern as for the usual semantic congruency effects were found (Iordanescu et al., 2008; Molholm et al., 2004; Tsilionis & Vatakis, 2016). Reaction times were faster for the congruent trials when being compared to all other conditions; the incongruent, white noise and no sound trials. It appears as if the congruent sounds also reveal a facilitating effect as they matched the visual stimulus, while being ambiguous. This resembles the effects of multisensory enhancement throughout working memory tasks, which enlarges when the stimuli are semantically congruent (Meredith, 2002). Throughout these specific paradigms, this effect was also described as the auditory enhancement effect, when a congruent auditory stimulus facilitates the visual search or recognition (Iordanescu et al., 2008; Maezawa et al., 2022). This was in line with our first hypothesis. In addition, also the third hypothesis was confirmed since the no sound condition was neutral compared to the incongruent condition. However, it did not confirm our second hypothesis as the incongruent condition did not reveal slower reaction times compared to the white noise trials. Because of this, the interference effect appears to be less likely causing these outcomes. As Maezawa et al. (2022) described, if the incongruent trials would interfere based on semantic incongruency, it would reveal slower reaction times compared to the white noise and no sound trials which these authors did find. It could be that due to the used stimuli, the white noise was perceived as an incongruent sound since sounds such as the noise of a blender were used, resulting in comparable reaction times. However, this remains a possible hypothesis. These results are, however, in line with those found by Iordanescu et al. (2008),

whom also found no differences in reaction times for distractor-related or incongruent and no sound trials. Throughout all of the possible sound conditions, the overall accuracies revealed no differences. This is not surprising, as for the accuracies the literature did not reach a consensus yet. During some studies, higher accuracies were found for semantic congruent trials compared to others (Chen & Spence, 2010; Molholm et al., 2004). However, in other studies no significant differences for the accuracies were found (Iordanescu et al., 2008; Suied et al., 2009). Overall, for the accuracies only the third hypothesis appears to be confirmed.

Throughout this study, a substantial perceptual learning effect appeared for both the reaction times and the accuracies because of the stimulus repetition (Furmanski & Engel, 2000; Karni & Sagi, 1991). When the visual stimulus was repeated, the recognition was facilitated as was reflected in faster reaction times and higher accuracies. It appears as if the task became easier, once participants became familiar with the visual stimuli (Karni & Sagi, 1991). Interestingly, the previously described study by Maezawa et al. (2022) performed the control experiments to account for learning effects, since before their first 3 experiments a familiarity task was performed. However, for these authors the pattern of the results remained similar, which some other studies appear to replicate (Zweig et al., 2015). A possible explanation could be the differences in the paradigms, as Maezawa et al. (2022) used a visual search task and we used a visual recognition task with ambiguous sounds. In addition, when investigating the overall means for the accuracies, no differences were found for congruent trials when comparing the first and second presentation. It could be that for the congruent trials the maximum accuracy was reached after the first presentation, which does remind us of a facilitating effect. Because of the perceptual learning effect and differences in sound conditions for the accuracies, the first and second presentation of the stimuli were investigated and interpreted separately.

As the first presentation of stimuli got investigated separately, only a significant difference for the accuracies was found. The results revealed a main significant effect for the sound conditions, where the congruent trials appeared to be responded to more accurately compared to the no sound trials. It also showed a trend where the congruent trials were responded to more accurately compared to the incongruent trials resembling the typical semantic congruency effect (Chen & Spence, 2010; Molholm et al., 2004; Tsilionis & Vatakis, 2016), however, this remained non-significant. This partially confirmed our first and third hypothesis, however, caution is necessary. These results can be rather difficult to interpret, as just a general auditory facilitation compared to the unisensory or no sound condition should also reflect on the white noise condition. Because of the significant

perceptual learning effect that was found, the results of the first presentation of the stimuli might be more reliable in comparison to the results of the repetition trials when trying to measure the semantic congruency effects.

For the second presentation of the stimuli, or the repetition, more interesting results were found. Focusing on the reaction times, a significant main effect for the sound conditions was found. Results revealed the congruent trials to be responded to faster in comparison to the white noise trials. This is somewhat in line with our first hypothesis and previous results (Maezawa et al., 2022), however, no significant differences in comparison to the incongruent or no sound trials were found. It does confirm our third hypothesis, but is inconsistent with our second hypothesis. Results also revealed a trend of semantic congruency (Iordanescu et al., 2008; Maezawa et al., 2022; Molholm et al., 2004; Suied et al., 2009; Tsilionis & Vatakis, 2016), where congruent trials were responded to faster in comparison with incongruent trials, however, this remained non-significant.

The accuracies for the second repetition also revealed a significant main effect of sound condition. Despite all our hypotheses and especially the third hypothesis, the no sound condition was responded to more accurately in comparison to all other sound conditions; congruent, incongruent and white noise. The literature has not reached a consensus for the accuracies, however, this is not consistent with any of the previous literature (Chen & Spence, 2010; Molholm et al., 2004). These results were rather unexpected and remain difficult to explain. It is possible to take into consideration that the congruent trials already appeared to reach maximum accuracy during the first presentation, as no differences in accuracy were found for this condition. It remains surprising as the no sound condition reveals the highest accuracies, as this could be seen as a unisensory presentation of information compared to the other multisensory conditions (Meredith, 2002). We do argue, however, that the second presentation or repetition of the visual stimuli reveals less reliable results because of the perceptual learning effect. Therefore, it could be considered less important.

There are a few important limitations to this study that must be acknowledged. As this paradigm was rather new, there is still room for improvement. First, as previously mentioned, the perceptual learning effect (Furmanski & Engel, 2000; Karni & Sagi, 1991) is possibly confounding the results during this study (Jones-Cage, 2017). During the second presentation of the visual stimuli, the reaction times and accuracies were significantly different. As for the accuracies, the sound condition did significantly change when comparing the first and second presentation, so the overall results can be less reliable. Second, when selecting the ambiguous sounds, sounds were used with a recognizability ranging from 0 to 40%. It is thus highly

plausible that different levels of ambiguity were present during this paradigm. This could be a possible confound, as different levels of ambiguity might reveal higher or lower semantic congruency effects but this remains a hypothesis.

Interesting approaches for future research did arose when conducting this study. As the semantic congruency effect appears somewhat for the reaction times but remains rather unclear, this paradigm could be tested using more recognizable and easier stimuli. Taking our limitations into account, different levels of ambiguity or recognizability could be tested. Ranging all the way from simple stimuli, such as a dog and the sound of barking, to more ambiguous stimuli, such as a fan and the sound of wind blowing, differences in the semantic congruency effects can be tested.

Conclusion

Overall, at first glance for the reaction times the semantic congruency effect appeared to be somewhat present, as congruent trials were responded to faster compared to all other conditions. It does reveal the typical impairment for incongruent trials, however, white noise and no sound trials revealed similar reaction times. For the accuracies there were no significant differences when comparing the four sound conditions. Caution is necessary, as a substantial perceptual learning effect was present. When focusing on the first presentation of the visual stimuli, the semantic congruency effect for reaction times disappeared. For the accuracies, a trend resembling the semantic congruency effect was present but remained non-significant. As the perceptual learning effect and different levels of ambiguity were present, this paradigm might benefit from important adjustments. Using more stimuli to avoid repetition and mapping the different levels of ambiguity could lead to interesting results and possible variations of the semantic congruency effect. However, for now, this remains a hypothesis.

References

Allaire, J. J. (n.d.). *RStudio: Integrated Development Environment for R*.

Andersen, R. A., & Cui, H. (2009). Intention, Action Planning, and Decision Making in Parietal-Frontal Circuits. *Neuron*, *63*(5), 568–583. https://doi.org/10.1016/j.neuron.2009.08.028

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal Representation of Space in the Posterior Parietal Cortex and Its Use in Planning Movements. *Annual Review of Neuroscience*, *20*(1), 303–330. https://doi.org/10.1146/annurev.neuro.20.1.303

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A. (1998). Working memory. *Comptes Rendus de l'Académie Des Sciences - Series III - Sciences de La Vie*, *321*(2), 167–173. https://doi.org/10.1016/S0764-4469(97)89817-4

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Botta, F., Santangelo, V., Raffone, A., Sanabria, D., Lupiáñez, J., & Belardinelli, M.

O. (2011). Multisensory integration affects visuo-spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1099–1109. https://doi.org/10.1037/a0023513

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K.-P., Zilles, K., & Fink, G. R. (2001). Polymodal Motion Processing in Posterior Parietal and Premotor Cortex: A Human fMRI Study Strongly Implies Equivalencies between Humans and Monkeys. *Neuron*, *29*(1), 287–296. https://doi.org/10.1016/S0896-6273(01)00198-2

Cappe, C., Rouiller, E. M., & Barone, P. (2012). *Cortical and Thalamic Pathways for Multisensory and Sensorimotor Interpla*. https://www.ncbi.nlm.nih.gov/books/NBK92866

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working Memory From the Psychological and Neurosciences Perspectives: A Review. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.00401

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. https://doi.org/10.1016/j.cognition.2009.10.012

Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews Neuroscience*, *3*(2), Article 2. https://doi.org/10.1038/nrn726

Cowan, N. (1995). *Attention and memory: An integrated framework* (pp. xv, 321). Oxford University Press.

Cowan, N. (1999). An Embedded-Processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

https://doi.org/10.1017/S0140525X01003922

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. https://doi.org/10.1162/jocn.2011.21625

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*(8), 1388–1404. https://doi.org/10.3758/BF03195006

Driver, J., & Noesselt, T. (2008). Multisensory Interplay Reveals Crossmodal Influences on 'Sensory-Specific' Brain Regions, Neural Responses, and Judgments. *Neuron*, *57*(1), 11–23. https://doi.org/10.1016/j.neuron.2007.12.013

Egeth, H., & Smith, E. E. (19670101). Perceptual selectivity in a visual recognition task. *Journal of Experimental Psychology*, *74*(4, Pt.1), 543. https://doi.org/10.1037/h0024774

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., & Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of Neurophysiology*, *76*(1), 141–157. https://doi.org/10.1152/jn.1996.76.1.141

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, *10*(1), 77–83. https://doi.org/10.1016/S0926-6410(00)00024-0

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport*, *16*(5), 419.

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition:

Object specificity and size invariance. *Vision Research*, *40*(5), 473–484. https://doi.org/10.1016/S0042-6989(99)00134-0

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*(6784), Article 6784. https://doi.org/10.1038/35012613

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

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. https://doi.org/10.1162/089892999563544

Goolkasian, P., & Foos, P. W. (2005). Bimodal Format Effects in Working Memory. *The American Journal of Psychology*, *118*(1), 61–78. https://doi.org/10.2307/30039043

Hanley, J. R., Young, A. W., & Pearson, N. A. (1991). Impairment of the visuo-spatial sketch pad. *The Quarterly Journal of Experimental Psychology Section A*, *43*(1), 101–125. https://doi.org/10.1080/14640749108401001

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, *15*(3), 548–554. https://doi.org/10.3758/PBR.15.3.548

James, T. W., & Stevenson, R. A. (2012). *The Use of fMRI to Assess Multisensory Integratio*. https://www.ncbi.nlm.nih.gov/books/NBK92856

Jones, E. G., & Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, *93*(4), 793–820. https://doi.org/10.1093/brain/93.4.793

Jones-Cage, C. (2017). Identifying confounding factors in psychology research. In *Activities for teaching statistics and research methods: A guide for psychology instructors*

(pp. 91–93). American Psychological Association. https://doi.org/10.1037/0000024-019

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*(11), 4966–4970. https://doi.org/10.1073/pnas.88.11.4966

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–384. https://doi.org/10.1016/j.actpsy.2010.03.010

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. https://doi.org/10.1007/s00221-004-1913-2

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., & Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Human Brain Mapping*, *19*(4), 213–223. https://doi.org/10.1002/hbm.10112

Linden, D. E. J. (2007). The Working Memory Networks of the Human Brain. *The Neuroscientist*, *13*(3), 257–267. https://doi.org/10.1177/1073858406298480

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP: Graphical Statistical Software for Common Statistical Designs. *Journal of Statistical Software*, *88*, 1–17. https://doi.org/10.18637/jss.v088.i02

Macaluso, E., & Driver, J. (2005). Temporal aspects of multisensory integration in the brain. *Trends in Neurosciences*, *5*(28), 264–271. https://doi.org/10.1016/j.tins.2005.03.008

Maezawa, T., Kiyosawa, M., & Kawahara, J. I. (2022). Auditory enhancement of visual searches for event scenes. *Attention, Perception, & Psychophysics*, *84*(2), 427–441. https://doi.org/10.3758/s13414-021-02433-8

Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., Van der Linden, M., Collette, F., Salmon, E., Seurinck, R., Fias, W., & Maquet, P. (2010). The Commonality of Neural Networks for Verbal and Visual Short-term Memory. *Journal of Cognitive Neuroscience*, *22*(11), 2570–2593. https://doi.org/10.1162/jocn.2009.21378

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J.-P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory Interactions within Human Primary Cortices Revealed by BOLD Dynamics. *Cerebral Cortex*, *17*(7), 1672–1679. https://doi.org/10.1093/cercor/bhl077

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research*, *14*(1), 31–40. https://doi.org/10.1016/S0926-6410(02)00059-9

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. https://doi.org/10.1152/jn.1986.56.3.640

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual–Auditory Object Recognition in Humans: A High-density Electrical Mapping Study. *Cerebral Cortex*, *14*(4), 452–465. https://doi.org/10.1093/cercor/bhh007

Moore, A. B., Li, Z., Tyner, C. E., Hu, X., & Crosson, B. (2013). Bilateral basal ganglia activity in verbal working memory. *Brain and Language*, *125*(3), 316–323. https://doi.org/10.1016/j.bandl.2012.05.003

Oliveri, M., Turriziani, P., Carlesimo, G. A., Koch, G., Tomaiuolo, F., Panella, M., & Caltagirone, C. (2001). Parieto-frontal Interactions in Visual-object and Visual-spatial Working Memory: Evidence from Transcranial Magnetic Stimulation. *Cerebral Cortex*, *11*(7), 606–618. https://doi.org/10.1093/cercor/11.7.606

Pahor, A., Collins, C., Smith-Peirce, R. N., Moon, A., Stavropoulos, T., Silva, I., Peng, E., Jaeggi, S. M., & Seitz, A. R. (2021). Multisensory Facilitation of Working Memory

Training. *Journal of Cognitive Enhancement, 5*(3), 386–395. https://doi.org/10.1007/s41465-020-00196-y

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, *6*(2), Article 2. https://doi.org/10.1038/nrn1603

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. https://doi.org/10.1016/j.neuroscience.2005.06.005

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*. https://www.frontiersin.org/article/10.3389/fnhum.2015.00197

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience, 22*(8), Article 8. https://doi.org/10.1038/s41593-019-0428-x

Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hearing Research*, *258*(1), 89–99. https://doi.org/10.1016/j.heares.2009.04.009

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. https://doi.org/10.1002/hbm.23918

Romanski, L. M., & Hwang, J. (2012). Timing of audiovisual inputs to the prefrontal cortex and multisensory integration. *Neuroscience, 214*, 36–48. https://doi.org/10.1016/j.neuroscience.2012.03.025

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology*, *15*(4), 454–458. https://doi.org/10.1016/j.conb.2005.06.008

Scimeca, J. M., Kiyonaga, A., & D'Esposito, M. (2018). Reaffirming the Sensory

Recruitment Account of Working Memory. *Trends in Cognitive Sciences*, *22*(3), 190–192. https://doi.org/10.1016/j.tics.2017.12.007

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), Article 4. https://doi.org/10.1038/nrn2331

Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research*, *194*(1), 91–102. https://doi.org/10.1007/s00221-008-1672-6

Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*. https://doi.org/10.3389/fnint.2015.00019

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Thelen, A., Cappe, C., & Murray, M. M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *NeuroImage*, *62*(3), 1478–1488. https://doi.org/10.1016/j.neuroimage.2012.05.027

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *48*(3), 380–398. https://doi.org/10.1037/1196-1961.48.3.380

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751–754. https://doi.org/10.1038/nature02466

Tsilionis, E., & Vatakis, A. (2016). Multisensory binding: Is the contribution of synchrony and semantic congruency obligatory? *Current Opinion in Behavioral Sciences*, *8*, 7–13. https://doi.org/10.1016/j.cobeha.2016.01.002

Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 151–161. https://doi.org/10.1016/S0022-5371(84)90104-X

Welch, R. B., & Warren, D. H. (19810101). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638. https://doi.org/10.1037/0033-2909.88.3.638

Xu, Y. (2018). Sensory Cortex Is Nonessential in Working Memory Storage. *Trends in Cognitive Sciences*, *22*(3), 192–193. https://doi.org/10.1016/j.tics.2017.12.008

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), Article 7080. https://doi.org/10.1038/nature04262

Yörük, H., Santacroce, L. A., & Tamber-Rosenau, B. J. (2020). Reevaluating the sensory recruitment model by manipulating crowding in visual working memory representations. *Psychonomic Bulletin & Review*, *27*(6), 1383–1396. https://doi.org/10.3758/s13423-020-01757-0

Zhao, Y.-J., Kay, K. N., Tian, Y., & Ku, Y. (2022). Sensory Recruitment Revisited: Ipsilateral V1 Involved in Visual Working Memory. *Cerebral Cortex*, *32*(7), 1470–1479. https://doi.org/10.1093/cercor/bhab300

Zweig, L. J., Suzuki, S., & Grabowecky, M. (2015). Learned face–voice pairings facilitate visual search. *Psychonomic Bulletin & Review*, *22*(2), 429–436. https://doi.org/10.3758/s13423-014-0685-3

Additional reference to an internship report:

Bettens, F. (2022, March). *Predictive Coding: the Influence of Contextual Information on Object Recognition.*

Appendix

Table 2

Repeated Measures ANOVA (RTs)

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Sound Condition	None	0.421	3.000	0.140	4.868	0.003	0.074
	Greenhouse-Geisser	0.421	2.942	0.143	4.868	0.003	0.074
Residuals	None	5.281	183.000	0.029			
	Greenhouse-Geisser	5.281	179.465	0.029			
Repetition	None	6.660	1.000	6.660	126.893	< .001	0.675
Residuals	None	3.202	61.000	0.052			
Sound Condition ✻ Repetition	None	0.059 [a]	3.000 [a]	0.020 [a]	0.662 [a]	0.577 [a]	0.011
	Greenhouse-Geisser	0.059	2.443	0.024	0.662	0.547	0.011
Residuals	None	5.468	183.000	0.030			
	Greenhouse-Geisser	5.468	149.009	0.037			

Note. Type III Sum of Squares using JASP.

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table 3

Post Hoc Comparisons - Sound Condition (RTs)

		Mean Difference	SE	t	p_{holm}
Congruent	Incongruent	-0.054	0.022	-2.516	0.051
	White Noise	-0.077	0.022	-3.577	0.003
	No Sound	-0.062	0.022	-2.888	0.022
Incongruent	White Noise	-0.023	0.022	-1.061	0.870
	No Sound	-0.008	0.022	-0.372	0.983
White Noise	No Sound	0.015	0.022	0.689	0.983

Note. Results are averaged over the levels of: Repetition

Table 4

Repeated Measures ANOVA (ACC)

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Sound Condition	None	0.022 [a]	3.000 [a]	0.007 [a]	1.569 [a]	0.198 [a]	0.025
	Greenhouse-Geisser	0.022	2.424	0.009	1.569	0.207	0.025
Residuals	None	0.871	183.000	0.005			

Repeated Measures ANOVA (ACC)

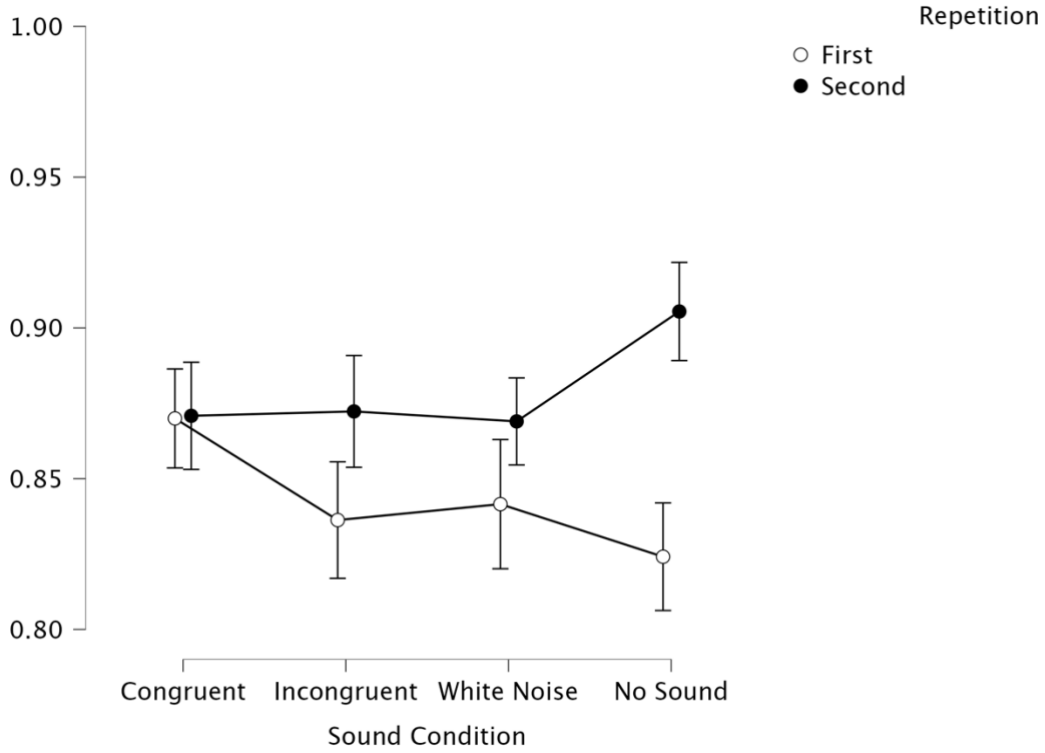
Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
	Greenhouse-Geisser	0.871	147.879	0.006			
Repetition	None	0.164	1.000	0.164	70.598	< .001	0.536
Residuals	None	0.142	61.000	0.002			
Sound Condition ✻ Repetition	None	0.104 [a]	3.000 [a]	0.035 [a]	5.779 [a]	< .001 [a]	0.087
	Greenhouse-Geisser	0.104	2.527	0.041	5.779	0.002	0.087
Residuals	None	1.100	183.000	0.006			
	Greenhouse-Geisser	1.100	154.131	0.007			

Note. Type III Sum of Squares using JASP.

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Figure 7

*Interaction Effect Sound Condition * Repetition (ACC)*



Note. Representation of the interaction effect Sound Condition * Repetition for the accuracies. The dots represent the means and the bars represent error bars.

Table 5

Post Hoc Comparisons - Sound Condition ✲ Repetition (ACC)

		Mean Difference	SE	t	p_{holm}
Congruent, First	Incongruent, First	0.034	0.013	2.558	0.161
	White Noise, First	0.028	0.013	2.157	0.349
	No Sound, First	0.046	0.013	3.480	0.012
	Congruent, Second	-8.706×10^{-4}	0.013	-0.068	1.000
	Incongruent, Second	-0.002	0.012	-0.194	1.000
	White Noise, Second	9.874×10^{-4}	0.012	0.082	1.000
	No Sound, Second	-0.035	0.012	-2.956	0.069
Incongruent, First	White Noise, First	-0.005	0.013	-0.401	1.000
	No Sound, First	0.012	0.013	0.922	1.000
	Congruent, Second	-0.035	0.012	-2.882	0.083
	Incongruent, Second	-0.036	0.013	-2.812	0.102
	White Noise, Second	-0.033	0.012	-2.727	0.113
	No Sound, Second	-0.069	0.012	-5.765	< .001
White Noise, First	No Sound, First	0.017	0.013	1.323	1.000
	Congruent, Second	-0.029	0.012	-2.441	0.181
	Incongruent, Second	-0.031	0.012	-2.563	0.161
	White Noise, Second	-0.027	0.013	-2.141	0.349
	No Sound, Second	-0.064	0.012	-5.325	< .001
No Sound, First	Congruent, Second	-0.047	0.012	-3.894	0.003
	Incongruent, Second	-0.048	0.012	-4.016	0.002
	White Noise, Second	-0.045	0.012	-3.739	0.005
	No Sound, Second	-0.081	0.013	-6.347	< .001
Congruent, Second	Incongruent, Second	-0.001	0.013	-0.111	1.000
	White Noise, Second	0.002	0.013	0.141	1.000
	No Sound, Second	-0.035	0.013	-2.626	0.144
Incongruent, Second	White Noise, Second	0.003	0.013	0.252	1.000
	No Sound, Second	-0.033	0.013	-2.515	0.161
White Noise, Second	No Sound, Second	-0.036	0.013	-2.767	0.107

Table 6

Repeated Measures ANOVA, first presentation only (ACC)

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Sound Condition	0.070	3	0.023	3.876	0.010	0.060
Residuals	1.104	183	0.006			

Note. Type III Sum of Squares using JASP.

Table 7

Post Hoc Comparisons - Sound Condition for the first presentation (ACC)

		Mean Difference	SE	t	p_{holm}
Congruent	Incongruent	0.034	0.014	2.416	0.083
	White Noise	0.028	0.014	2.037	0.172
	No Sound	0.046	0.014	3.287	0.007
Incongruent	White Noise	-0.005	0.014	-0.379	0.770
	No Sound	0.012	0.014	0.871	0.770
White Noise	No Sound	0.017	0.014	1.250	0.639

Table 8

Repeated Measures ANOVA, second presentation only (RTs)

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Sound Condition	0.256	3	0.085	3.777	0.012	0.058
Residuals	4.130	183	0.023			

Note. Type III Sum of Squares

Table 9

Post Hoc Comparisons - Sound Condition for the second presentation (RTs)

		Mean Difference	SE	t	p_{holm}
Congruent	Incongruent	-0.065	0.027	-2.427	0.081
	White Noise	-0.087	0.027	-3.225	0.009
	No Sound	-0.046	0.027	-1.710	0.356
Incongruent	White Noise	-0.022	0.027	-0.798	0.852
	No Sound	0.019	0.027	0.717	0.852
White Noise	No Sound	0.041	0.027	1.515	0.394

Table 10

Repeated Measures ANOVA, second presentation only (ACC)

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Sound Condition	0.056	3	0.019	3.974	0.009	0.061
Residuals	0.866	183	0.005			

Note. Type III Sum of Squares using JASP.

Table 11

Post Hoc Comparisons - Sound Condition for the second presentation (ACC)

		Mean Difference	SE	t	p_{holm}
Congruent	Incongruent	-0.001	0.012	-0.118	1.000
	White Noise	0.002	0.012	0.150	1.000
	No Sound	-0.035	0.012	-2.800	0.028
Incongruent	White Noise	0.003	0.012	0.269	1.000
	No Sound	-0.033	0.012	-2.682	0.032
White Noise	No Sound	-0.036	0.012	-2.950	0.022