# SPEED PERCEPTION IN MULTISENSORY PROFILES: WHAT IS THE EFFECT OF ATTENTION?

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF THE MIDDLE
EAST TECHNICAL UNIVERSITY

BY

EFSUN KAVAKLIOĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

IN
THE DEPARTMENT OF COGNITIVE SCIENCE

JANUARY 2022

# SPEED PERCEPTION IN MULTISENSORY PROFILES: WHAT IS THE EFFECT OF ATTENTION?

Submitted by EFSUN KAVAKLIOĞLU in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics**

Dr. Ceyhan Temürcü
Head of Department, **Cognitive Science**

Assist. Prof. Dr. Umut Özge
Supervisor, **Cognitive Science Dept., METU**

Assist. Prof. Dr. Hulusi Kafalıgönül
Co-Supervisor, **UMRAM and Neuroscience Dept., Bilkent U.**

**Examining Committee Members:**

Assist. Prof. Dr. Murat Perit Çakır
Cognitive Science Dept., METU

Assist. Prof. Dr. Umut Özge
Supervisor, Cognitive Science Dept., METU

Assist. Prof. Dr. Hacı Hulusi Kafalıgönül
Neuroscience Dept., Bilkent University

Assist. Prof. Dr. Dicle Dövencioğlu
Psychology Department, METU

Assist. Prof. Dr. Funda Yıldırım
Computer Engineering Department, Yeditepe University

Date: 07-02-2022

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

**Name, Last Name: Efsun Kavaklıoğlu**

**Signature: ____________________**

# ABSTRACT

SPEED PERCEPTION IN MULTISENSORY PROFILES: WHAT IS THE EFFECT OF ATTENTION?

Kavaklıoğlu, Efsun

M.Sc, Department of Cognitive Science

Supervisor: Assist. Prof. Dr. Umut Özge

Co-Supervisor: Assist. Prof. Dr. Hulusi Kafalıgönül

January 2022, 70 pages

Multisensory processing and crossmodal interactions in the temporal domain are crucial for survival in a dynamic environment. Temporal ventriloquism illusion demonstrates the importance of the crossmodal interactions in the temporal domain and the influences of the auditory signals (e.g., auditory time intervals) on visual perception. Attention is another mechanism playing a critical role in sensory processing, and it allows us to prioritize relevant information in the visual field. Previous studies have shown that attention can modulate multisensory processing and crossmodal interactions. The current thesis was focused on understanding whether spatial attention modulates audiovisual interactions in the temporal domain. An apparent motion paradigm of temporal ventriloquism illusion was used. The experimental design was based on auditory time interval effects on perceived visual speed. Besides speed judgment, participants performed a secondary task on discriminating static clicks. Using endogenous cues, we manipulated attention in the auditory domain and systematically changed the difficulty level of the secondary task. The results revealed a significant effect of task difficulty on audiovisual interactions in the temporal domain, such that the effects of auditory time intervals on perceived speed became larger with an increasing level of task difficulty. Moreover, the cueing effect became significant when the difficulty level of the secondary task was low, suggesting that the participants might have more resources available to process informative cues. Future studies aimed to determine the effects of task difficulty on attention and audiovisual processes may provide further insights on the involvement of attention in multisensory processing.

Key Words: attention, audiovisual interactions, speed discrimination, motion, multisensory

# ÖZ

ÇOKLU DUYUSAL PROFİLLERDE HIZ ALGILAMASI: DİKKATİN ETKİSİ NEDİR?

Kavaklıoğlu, Efsun

Yüksek Lisans, Bilişsel Bilimler bölümü

Tez Yöneticisi: Dr. Öğr. Üyesi Umut Özge

Ortak Tez Yöneticisi: Dr. Öğr. Üyesi Hulusi Kafalıgönül

Ocak 2022, 70 sayfa

Zamansal alandaki çoklu-duyulu işleme ve çapraz modaliteli etkileşimler, dinamik bir ortamda hayatta kalmak için çok önemlidir. Zamansal ventrilok ilüzyonu, zamansal alandaki çapraz modalite etkileşimlerinin önemini ve işitsel sinyallerin (örneğin, işitsel zaman aralıkları) görsel algı üzerindeki etkilerini gösterir. Dikkat, duyusal işlemede kritik rol oynayan bir başka mekanizmadır ve görsel alanda ilgili bilgilere öncelik vermemizi sağlar. Önceki çalışmalar, dikkatin çoklu-duyulu işlemeyi ve çapraz modaliteli etkileşimleri modüle edebileceğini göstermiştir. Mevcut tez, uzamsal dikkatin zamansal alandaki görsel-işitsel etkileşimleri modüle edip etmediğini anlamaya odaklanmıştır. Deneyde zamansal ventrilok illüzyonunun belirgin bir hareket paradigması kullanılmıştır. Deney tasarımı, algılanan görsel hız üzerindeki işitsel zaman aralığı etkilerine dayanmaktadır. Hız değerlendirmesinin yanı sıra, katılımcılar statik tıklamaları ayırt etme konusunda ikincil bir görev gerçekleştirmişlerdir. İpuçları kullanarak, işitsel alanda dikkat manipüle edilmiş ve ikincil görevin zorluk seviyesi sistematik olarak değiştirilmiştir. Sonuçlar, görev zorluğunun zamansal alandaki görsel-işitsel etkileşimler üzerinde önemli bir etkisini ortaya çıkarmıştır, öyle ki işitsel zaman aralıklarının algılanan hız üzerindeki etkileri, artan görev zorluğu seviyesi ile daha büyük hale gelmiştir. Ayrıca, ikincil görevin zorluk seviyesi düşük olduğunda ipucu etkisi anlamlı hale gelmiş, bu da katılımcıların bilgilendirici ipuçlarını işlemek için daha fazla kaynağa sahip olabileceğini düşündürmüştür. Görev zorluğunun dikkat ve görsel-işitsel süreçler üzerindeki etkilerini belirlemeyi amaçlayan gelecekteki çalışmalar, dikkatin çok-duyulu işlemeye dahil edilmesi hakkında daha fazla bilgi sağlayabilir.

Anahtar Sözcükler: dikkat, görsel-işitsel etkileşimler, hız ayrımı, hareket, çoklu duyusal

Right now, right here, surpass your limits

# ACKNOWLEDGEMENTS

First of all, I want to thank my thesis advisor Assistant Professor H. Hulusi Kafalıgönül for his great help, feedback and support. Prof. Kafalıgönül -with his intelligence, academic integrity and humor- is one of the best people Bilkent has introduced into my life. I am grateful for his determination, patience and encouragement. I would also like to thank my other advisor Assistant Professor Umut Özge for his help and encouragement when I needed it the most.

I would like to thank all of my lab-mates Dr. Şeyma Koç Yılmaz, İrem Akdoğan, Esra Nur Çatak, Afife Türker, Ayşenur Karaduman and Simay Üner who made coming to the lab everyday a great joy. Additionally I am grateful to Tuğçe Çabuk, Merve Kınıklıoğlu and Didenur Şahin Çevik for their support and friendship. I want to thank all these friends for their feedback, input and support. Also, I want to thank Aysun Duyar for her help and feedback during the experimental design of the study. She was great to work with and I am thankful for her input.

I literally couldn't be here without my mother Özlem Kavaklıoğlu. Her unconditional love and support drive me to try and be better every day. She deserves all the credit because she is the best mom I could imagine. I also want to thank my late father Fatih Kavaklıoğlu, for being my best friend and the best dad ever.

I want to thank my dear friends Ayça Güven, Aytuna Barkçin, Gökçe Yumuşak, Elif Bora, Atahan Güngör, Ömer Ekmekcioglu, Emre Mumcuoğlu, Mert Albaba and Deniz Akkaya for always being there for me when I need a friend to talk to.

I would also like to acknowledge the financial support provided by the Turkish Academy of Sciences (TÜBA-GEBİP Award Project).

# TABLE OF CONTENTS

CHAPTER

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **ANOVA** | Analysis Of Variance |
| **EEG** | Electroencephalography |
| **ERP** | Event-related Potentials |
| **FC** | Frontal Cortex |
| **IPS** | Intraparietal Sulcus |
| **ISI** | Interstimulus Interval |
| **LED** | Light Emitting Diode |
| **MS** | Millisecond |
| **MSI** | Multisensory Integration |
| **MSP** | Multisensory Processing |
| **MT** | Middle Temporal Area |
| **N/A** | Not Applicable |
| **PFC** | Prefrontal Cortex |
| **PFR** | Percentage of Fast Speed Responses |
| **PTV** | Percentage of Temporal Ventriloquism |
| **SC** | Superior Colliculus |
| **SD** | Standard Deviation |
| **SOA** | Stimulus Onset Asynchrony |
| **STS** | Superior Temporal Sulcus |
| **TWI** | Temporal Window of Integration |

# CHAPTER 1

## 1 INTRODUCTION

Every day, organisms rely on their senses to navigate through the world. Through their sensory organs, they gather spatial and temporal information about the external world; for instance, they get visual information from their eyes, auditory information from their ears, and olfactory information from their noses. For performing complex human behavior, accurate spatial and temporal information is necessary (Loeffler et al. 2018). Spatial information is generally related to the position, size, or area of objects whereas temporal information is about time. To understand what the spatial and temporal information from various modalities mean, organisms need to represent and process them appropriately (Postma et al., 2017). Accurate representation of external information is needed because organisms need to try to produce an appropriate response according to the situation they are in.

While we can process information unimodally, the human brain is also capable of processing multiple modalities at the same time as well, this process is called multisensory processing (MSP). In multimodal processing, the brain integrates the spatial and temporal information provided by the external environment, which is called multisensory integration (MSI). Multisensory integration is used in many different ways. For instance, it may be used in the evaluation of the perception of food qualities by the integration of olfactory and gustatory stimuli (de Araujo, Simon, 2009), or when associated with visual stimuli, it can be used in the localization of auditory stimuli as in ventriloquism effect (Thurlow, Jack, 1973).

Audiovisual ventriloquism paradigms notably provide great examples for spatiotemporal processing and integration. Hence they are commonly used to study MSP and MSI (Slutsky, Recanzone, 2001; Morein-Zamir et. al., 2003; Vroomen, Gelder, 2004). The spatial ventriloquism effect demonstrates the influences of visual information on auditory information, while the temporal ventriloquism displays the influences of auditory information on visual information (Bertelson, Aschersleben, 1998). These effects are mainly because vision has a superior spatial resolution whereas auditory processing has a higher temporal resolution (Loeffler et. al., 2018).

Audiovisual perception and ventriloquism effects have often been studied with dynamic processes such as motion. Apparent motion (Ramachandran, Anstis, 1986; Yantis, Nakama, 1998) is a phenomenon that occurs when stimuli with different timing and location are perceived as a single object that moves from one point to another and it is often used in studies examining audiovisual perception and ventriloquism effects (For examples see: Freeman, Driver, 2008; Ogulmus, Karacaoglu, Kafaligonul, 2018; Bruns, Getzmann, 2008). For example, researchers have shown that auditory stimuli can affect the perceived temporal properties of visual events and hence different aspects of apparent motion perception (Kafaligonul, Stoner, 2010; Kafaligonul, Oluk, 2015).

Unimodal and multimodal sensory perception and processing are necessary tools for performing both simple and complex tasks. However, we are constantly bombarded with stimuli from multiple resources and modalities. Since the processing capacity of humans is not infinite, there should be selection mechanisms to identify relevant information. For this reason, we routinely use a systematic process to prioritize the information according to relevance and importance. This selection process and mechanism is called *attention*.

Since attention and multisensory integration have important implications for daily life situations, there should be some form of relationship between these mechanisms. Researchers have been investigating the relationship between attention and multisensory integration using different modalities and paradigms. Especially the relationship of attention and the integration of audition and vision modalities have been studied using various illusions such as temporal ventriloquism.

## 1.1 Specific Aims and Research Questions

The general aim of this thesis is to investigate the relationship between attention and multisensory interactions. More specifically, this thesis focuses on understanding the relationship between temporal ventriloquism effect and attention by investigating the effect of endogenous cues on the audiovisual interaction of apparent motions and auditory stimuli with different time intervals.

In this thesis, two main hypotheses were investigated. First, we wanted to find whether or not attention affected the audiovisual interactions in the temporal domain, with a specific question of whether endogenous visual cues affected the level of integration of the auditory and visual stimuli. Second, we wanted to further understand whether task difficulty had an impact on endogenous attention and the level of audiovisual interactions. To determine the relationship between audiovisual interactions and attention, we designed a behavioral experiment and tested these hypotheses using a temporal ventriloquism paradigm based on apparent motions and auditory clicks with different time intervals. In order to evaluate perception and performance, we used *Response Times* (RT) and *Percentage of Temporal Ventriloquism (PTV)* as behavioral measures.

## 1.2 Organization of the Dissertation

This thesis is separated into distinct chapters for the convenience of the readers. The first chapter provides a short introduction to the general topic of the thesis and it also includes the research questions and the organizational information of the whole text. Chapter 2 is reserved for the literature review of the topics that are crucial in order to understand the research questions, the experiment design and the results. This chapter reviews the literature and includes information on the topics of multisensory interactions, attention, and the application of attention in multisensory research. In Chapter 3, the experiment is explained comprehensively. Chapter 4 shows the behavioral results. Chapter 5 is the section where the results and the future directions are discussed and the conclusion is stated. All the cited studies were listed in the References section.

# CHAPTER 2

# 2 LITERATURE REVIEW

## 2.1 Multisensory Processing

Sensory systems are vitally important in human life. For instance, vision and audition are crucial senses that inform us about possible threats and dangers of the outside world. All sensory systems bring unique and substantial information about the external environment. However, in most cases the stimulation from the external environment simultaneously triggers more than one sensory organ. The operation of processing information from an individual modality that interacts with and gets influenced by the information from another modality is called *Multisensory Integration* (Talsma, Senkowski, Soto-Faraco, 2010). Multisensory inputs are bound together when they are temporally and spatially close and the likelihood of them being bound together decreases as the temporal and/or spatial distance increases. The integration in the temporal domain can occur within a limited time window called the *Temporal Window of Integration (TWI)* (Donohue, Green, Woldorff, 2015). Figure 2.1 shows the relationship between the integration proportion of auditory and visual stimuli and Stimulus Onset Asynchrony (SOA). The proportion of audiovisual integration increases as the auditory and visual stimuli are temporally closer to each other.

Multisensory integration, and hence interaction across modalities are identified when the responses to stimuli from multiple modalities differ significantly from the responses to any of those modalities presented alone and this difference can not be explained by statistical summation effects or changed diligence (van Opstal, 2016). This process is an indispensable component of the human sensory experience, and thus it is usually effortless, smooth and automatic. (Angelaki, Dora E et al., 2009). Multisensory integration plays a crucial role in understanding whether different types of sensory input belong to the same object or not. Hence, it helps reduce the noise in our perceptual system by combining several inputs from different sensory modalities (Koelewijn, Bronkhorst, Theeuwes, 2010). Multisensory phenomena are topics of interest also because through them, it might be possible for us to understand the underlying cortical processes that create and maintain a coherent internal representation of the outside environment (Bertelson et al., 2000).

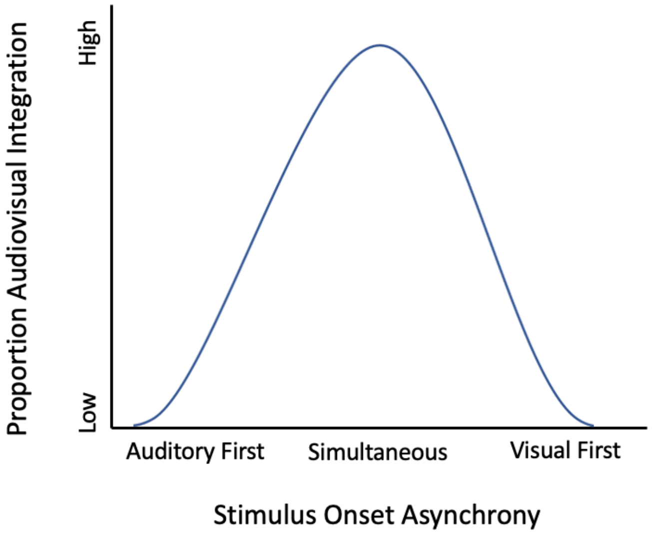


Figure 2.1 Temporal Window of Integration. As the temporal asynchrony between the auditory and visual stimuli decreases, the stimuli from the two modalities are more likely to be integrated.

"*Modality Appropriateness Theory*" suggests that the modality that is best suitable for the sensory task at hand is the dominant one during the integration process (Welch, Warren, 1980). Therefore, if the task requires precision of time rather than spatial precision, then the modality with a higher temporal resolution will be dominant over the modality with a less temporal resolution. Likewise, if the task requires spatial precision (rather than time), then the modality with a higher spatial resolution will be dominant. For instance, vision dominates audition in the spatial domain whereas audition is superior in the temporal domain.

Traditionally cross-modal interactions are thought to be automatic, however recently it has been shown that these interactions can also be modulated by attention (Santangelo, Macaluso, 2012). Using different paradigms, the relationship between multisensory processing and attention has been investigated. In the following subsections, we introduce the paradigms studied in the current study and also cover studies of attention within this context.

## 2.2 Audiovisual Interactions and Processing

Since we are mainly exposed to visual and auditory stimulation in life, the interactions between these modalities are extensively studied. For instance, the integration of auditory and visual stimuli is crucial in speech processing and it allows us to associate the sounds and mouth movements as coming from a single source rather than coming from multiple different ones, which ultimately leads to coherent speech comprehension (Besle et al., 2004).

Audiovisual interactions have been demonstrated by many different paradigms, wherein some of them the visual stimulus influences the auditory stimulus, and in others the auditory stimulus influences the visual stimulus. As discussed above, since audition has a superior temporal resolution, audition dominates over vision in the final percept of tasks that require temporal precision. On the other hand, if there is a spatial task, then vision dominates over audition in sensory processing. The well-known paradigms and illusions of audiovisual interactions (e.g., *Temporal Ventriloquism, McGurk Effect, Sound-Induced Flash Illusion and Streaming/ Bouncing Effect*) are briefly discussed below.

### *2.2.1 Temporal Ventriloquism*

When visual and auditory stimulation is presented in close temporal proximity, the sensory system mainly relies on the temporal information provided by audition. This well-known illusion in which auditory information dominates over visual information in the temporal domain is called "temporal ventriloquism". For instance, in Morein- Zamir, Soto-Faraco, and Kingstone (2003), where the task was a temporal order judgement of two brief visual flashes, the performance of participants increased

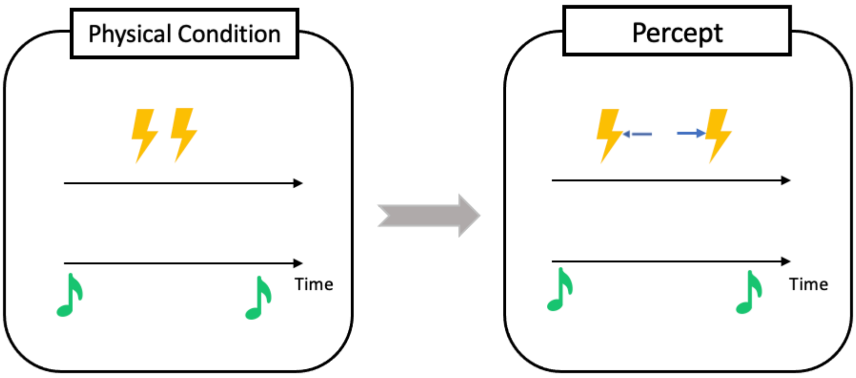


Figure 2.2 Representation of temporal ventriloquism illusion. The apparent onset of the visual stimuli, brief flash by the LEDs, shifts towards the onset of auditory stimuli. Hence, the final percept is typically different than the physical time interval.

when brief static clicks were introduced with longer time interval (ISI). This change in performance has been interpreted and described as brief sounds capturing the visual flashes in the temporal domain, increasing the perceived time interval between the visual events and ultimately leading to a performance increase in temporal order judgement of these visual events. A visual representation of the temporal ventriloquism effect can be found in Figure 2.2.

### *2.2.2 McGurk Effect*

One of the most known audiovisual interactions where a visual stimulus dominates the auditory stimulus is the McGurk Effect. McGurk effect happens when the visual lip movements of an individual articulating a syllable do not correspond to what that individual is audibly pronouncing. When there is such conflict, the sensory system typically relies on the information provided by the visual modality, and the observer perceives a different sound than the one being articulated (McGurk, MacDonald, 1976). Notably, it is a striking illusion by demonstrating the strong influences of audiovisual interactions and binding in the spatial domain on the final percept (Tiippana, 2014). An illustration of the McGurk effect can be seen in Figure 2.3. In this effect, the participant views the mouth of someone forming the syllable "Ga" and simultaneously hears the sound "Ba". In the final percept, the participant reports their experience as hearing the syllable "Da".

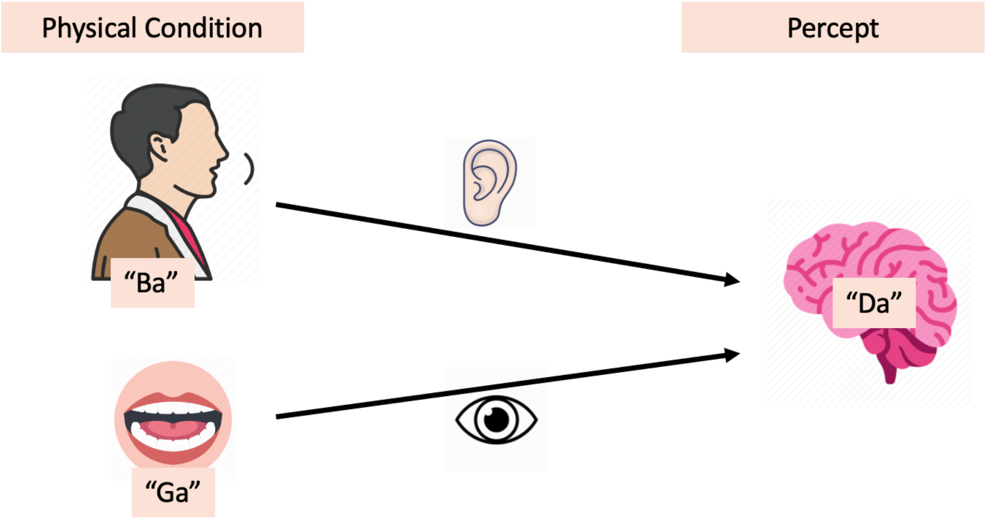


Figure 2.3 Illustration of the McGurk effect. In the physical condition, the participant sees the mouth of a person forming the syllable "Ga". Simultaneously, the sound of a person articulating "Ba" is played. As a result, the participant perceives a person sounding the syllable "Da".

### *2.2.3 Sound-Induced Flash Illusion*

In this illusion, a single flash is presented in the periphery with multiple auditory clicks. When the number of clicks is more than one, the observers typically report seeing more than one visual flash (Rosenthal, Shimojo, Shams, 2009). In fact, this illusion has been considered as another powerful illusion demonstrating strong influences of auditory clicks on the final percept in the temporal domain (Hirst et al., 2020, see also Figure 2.4).

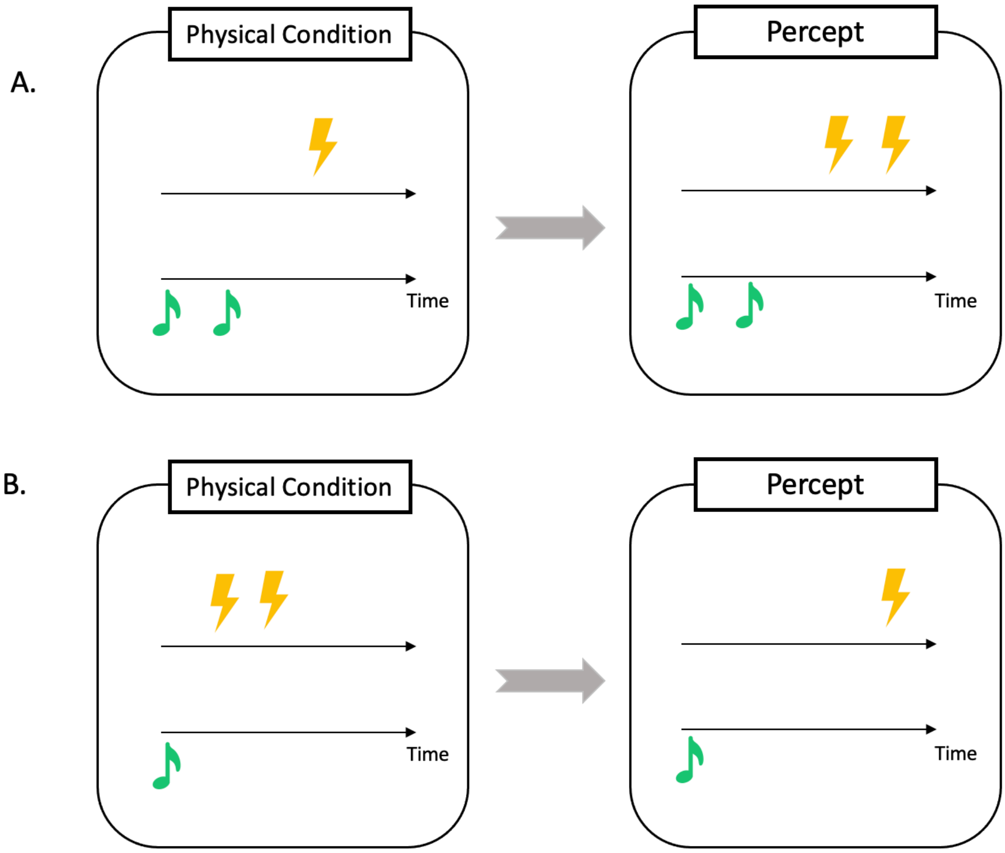


Figure 2.4 Illustration of two types of the sound-induced flash illusion. (A) In the physical condition, two clicks are followed by a single flash. However, even though there is a single physical flash, the participant perceives two separate flashes. (B) In this type of sound-induced flash illusion, a single click is followed by two consequent flashes. Similar to the percept in (A), the participant perceives a single flash even though there are two physical flashes.

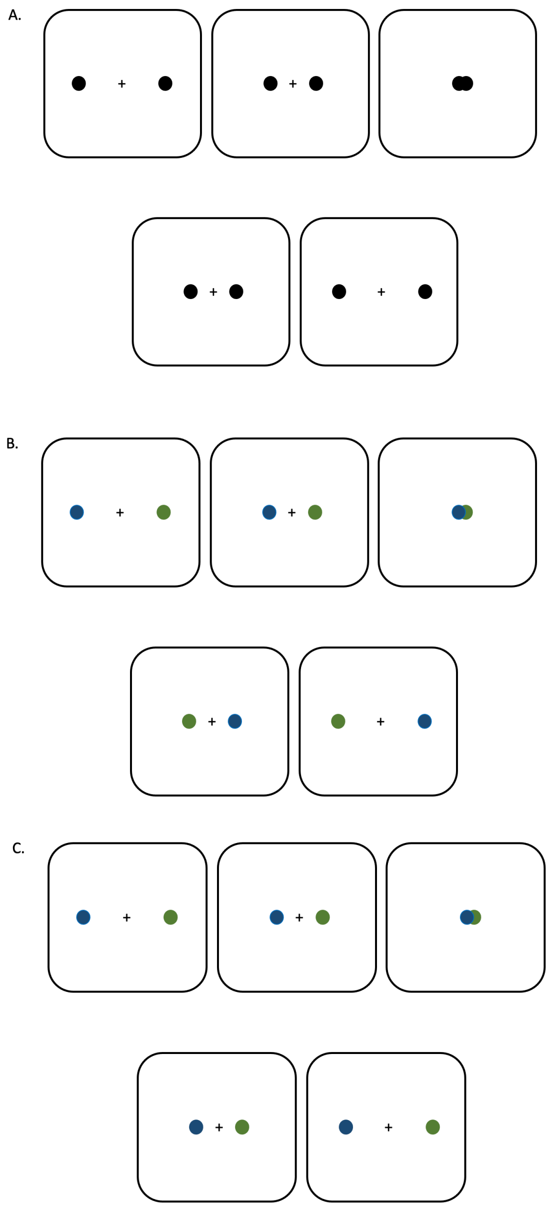


Figure 2.5 Illustration of the streaming/bouncing effect. (A) The physical conditions of a typical experiment based on this illusion are shown. Two circles on the right and left of the fixation point move towards each other, coincide and then move towards their starting points. (B) When a bouncy/springy sound is played simultaneously with the physical condition explained in (A), participants perceive it as if two circles move toward each other, coincide and bounce back to their starting points. (C) When no sound is played during the physical condition explained in (A), participants perceive it as if the two circles move towards each other, coincide and then stream through each other to end their motion in the position symmetrical to their starting points.

### 2.2.4 Streaming/Bouncing Effect

Another well-known audiovisual paradigm where auditory timing affects visual perception is called the "*Streaming/Bouncing Effect*". In this illusion, the timing of a brief auditory event (e.g., a click) alters the perception of two identical objects moving toward each other. The objects perfectly overlap and continue going towards the opposite side of where they came from. Notably, a silent motion is mostly perceived as 'streaming motion' but when an auditory stimulus is played at the moment of collision the motion is mostly perceived as a 'bouncing motion' (Massimo, Casco, 2009; Remijn et al., 2004). Figure 2.5 illustrates the streaming/bouncing effect where (A) shows the physical conditions of the paradigm, (B) shows the percept of participants when the physical conditions are presented simultaneously with a 'bouncy sound' and (C) shows the percept when no sound is played when the physical conditions in (A) are presented.

### 2.2.5 Audiovisual Studies

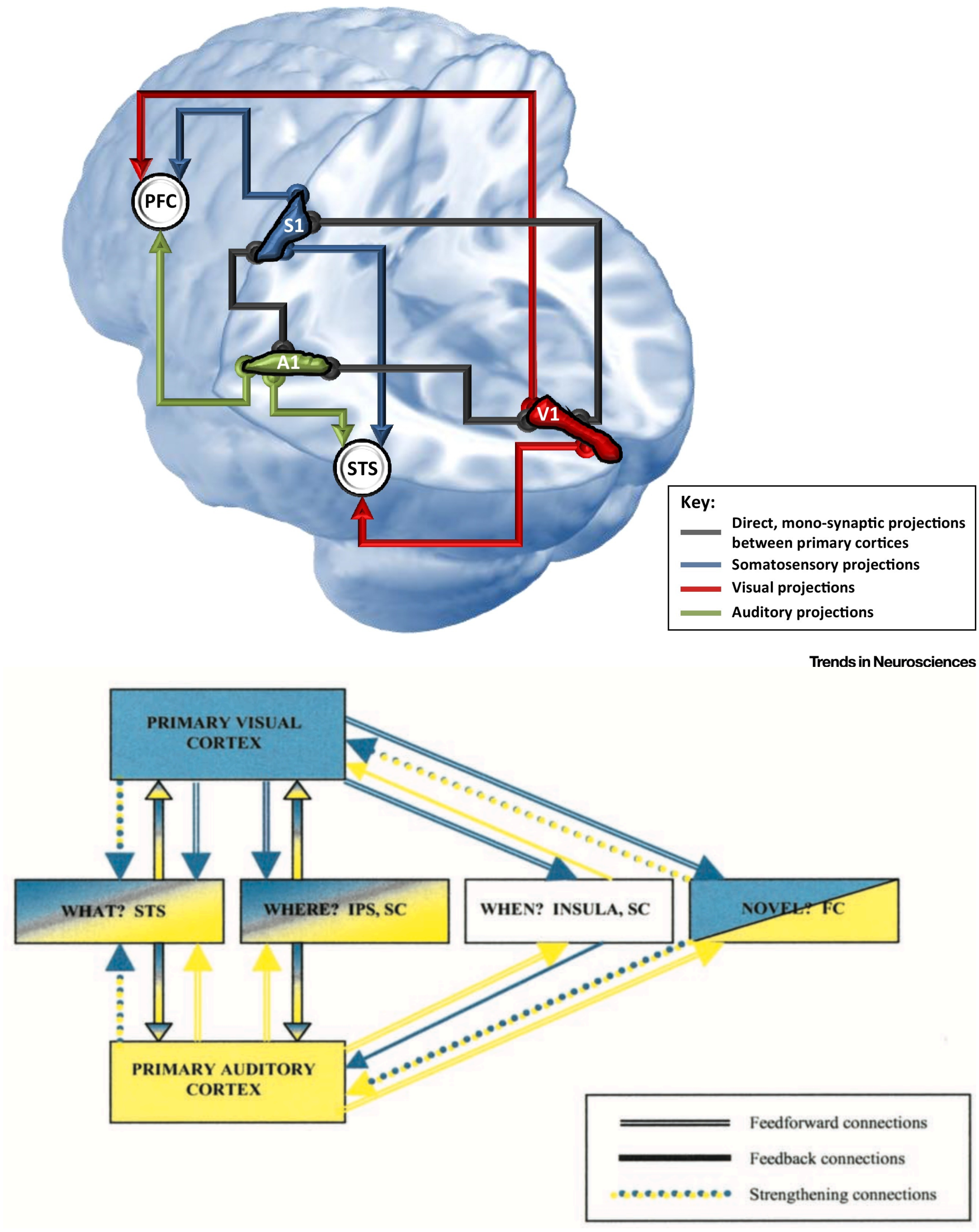
As it was mentioned previously, audiovisual interactions have been a topic of interest for a long time. Using the paradigms that were listed previously, researchers have tried to shed more light on the behavioral and neural correlates of audiovisual interactions. Many researchers have used the effects mentioned in this chapter and some other effects to understand the underlying mechanisms of audiovisual interactions.

Morein-Zamir et al. (2003) designed four experiments to see whether task-irrelevant sounds influenced a visual temporal order judgement task. In a visual temporal judgement task, participants are asked to report which visual stimulus was shown first. The experiment revealed that providing an auditory stimulus before the presentation of the first visual stimulus and after the presentation of the second visual stimulus improved participant performance as if auditory stimuli were increasing the temporal distance between the visual stimuli. Contrary to this effect, when the auditory stimuli were presented after the first visual stimulus was presented and before the second visual stimulus was presented, participant performance decreased as if the auditory stimuli decreased the temporal distance between the visual stimuli. The results of their experiments demonstrated that task-irrelevant auditory stimuli could modulate participant performance on visual temporal order judgement tasks. Freeman and Driver (2008) investigated the effect of the timing of static sounds on spatio-temporal processing of visual apparent motion generated by using visual bars. In the experiment, they used apparent motions with equal time intervals but they used auditory stimuli that slightly lagged the right flash and lead the left flash or slightly lagged the left flash and lead the right flash. As a result, they found out that the timing of the static sounds that slightly lagged or lead the apparent motion robustly influenced the perception of the apparent motion even though the auditory stimuli provided no directional information. Rather than focusing on the effects of the auditory stimuli on visual stimuli, Ogulmus, Karacaoglu and Kafaligonul (2018) used the temporal ventriloquism paradigm to investigate how audiovisual interactions are modulated by the spatial grouping mechanisms of the visual modality. In their experiments, they manipulated spatial grouping mechanisms such as proximity, uniform connectedness and similarity, and the participants were tasked to compare the speed of apparent motions with various auditory timing conditions. Through their experiments they found out that, the effect that auditory timing has on perceived speed varies for different spatial configurations and it can be modulated by spatial grouping principles.

## 2.3 Cortical Mechanisms of Audiovisual Integration

Human perceptual and cognitive systems have adapted to be operational in a multisensory environment. Although previous studies have suggested that multisensory integration was only linked to late-level cortex regions, recent studies offer a broader perspective. Many studies have supported the importance of early-level interactions between the senses in addition to late-level interactions. These interactions have also been deemed important since some perceptual illusions can be explained with only early-level interactions.

Overall, the broader perspective that was mentioned earlier proposes that early-level and late-level cognitive processes interact during multisensory integration. These interactions are thought to be accomplished through feedforward, feedback and recursive connections between various cortical and subcortical regions. The integration of crossmodal inputs seems to be the result of networks of brain areas rather than individual sites, while the individual components seem to be specialized for generating various aspects of cross-modal information (Calvert, 2001). Data from previous studies indicate that the active role of the regions and connections change by factors such as context, behavioral purpose, task and compatibility between stimuli from different modalities. Hence, the idea that these mechanisms are adaptable and flexible in multisensory integration processes is supported (Talsma, 2015; Van Atteveldt et al., 2014). For instance factors such as spatial and temporal proximity between stimuli, congruence and stimulus intensity have been shown to affect multisensory integration and processing. Moreover, many studies argue that the behavioral purpose of the task and paradigm affect the extent to which late cortical areas and feedback connections are involved in the integration process. The projections in multisensory integration between some early and late cortical regions can be seen in Figure 2.6. The projections shown in the figure are between the primary auditory cortex, primary visual cortex, prefrontal cortex and posterior



Trends in Neurosciences

Figure 2.6 This figure displays the projections between cortical areas involved in binding visual and auditory information. In the figure, A1 represents the primary auditory cortex, V1 the primary visual cortex, S1 primary somatosensory cortex, PFC the prefrontal cortex and STS the superior temporal sulcus, IPS the intraparietal sulcus, SC the superior colliculus and FC the frontal cortex (Adapted from Murray et. al, 2016).

superior temporal sulcus (STS). As it can be seen, information is projected to different regions for various reasons. In order to understand the content of the information, it is projected to STS while spatial information is projected to IPS and SC and temporal information is projected to SC and insula. For novel information, FC is prioritized as the region of projection.

The main anatomical regions that are crucial in multisensory processes can be seen in Figure 2.6. A1 is the Primary Auditory Cortex which is the first area in the temporal lobes of the brain and it is responsible for the processing of acoustic information. Similarly, V1 is the Primary Visual Cortex and it is responsible for the first stage of cortical processing of visual information. After processing the sensory information, A1 and V1 then project their outputs to different sites in order to process the information further. Some regions with emphasized roles in cross-modal integration that A1 and V1 project their output to can be listed as Superior Colliculus (SC), Superior Temporal Sulcus (STS), Inteparietal Sulcus (IPS) and Frontal Cortex (FC). Figure 2.5 B. displays the projections between the primary cortices and the cross-modal regions.

Many researchers have worked on localizing the neural activity of multisensory processes. Molholm et al. (2002) investigated the timing and topography of audiovisual interactions cortically using event-related potentials (ERPs) using a simple reaction time task. In the experiment, they presented isolated auditory stimuli, isolated visual stimuli and lastly, simultaneous auditory and visual stimuli. They compared the ERPs recorded during simultaneous stimuli presentation with the summation of ERPs in visual only and auditory only stimuli trials. They made this comparison to see whether the summation of the ERPs achieved from the solitary stimuli in different modalities was equivalent to the ERPs obtained from the simultaneous presentation of the different modalities. Their study displayed audiovisual interactions of the same polarity over right parieto-occipital, occipito-temporal, fronto-central and central scalp. The results suggested that audiovisual interactions could affect early visual processing. Moreover, the effect of audiovisual interactions over the scalp indicated both the modulation of unisensory activity and activity exclusive to multisensory processing. Mishra et al. (2007) investigated the early cross-modal interactions of audition and vision utilizing the sound induced flash illusion. In this study, they examined the timing and localization of the cortical processes that make this illusory effect possible using whole head ERP recordings. They found neural evidence that the illusion is made possible by the interaction of the visual, auditory and superior temporal cortex. According to the results, an enhanced cross-modal interaction in the auditory cortex might be responsible for triggering the perception of the illusion, and the illusion is produced by the interplay between the modality specific auditory and visual cortices and superior temporal cortex. Taking inspiration from the study with long-range visual motion that engaged with high-order cortical areas by Freeman and Driver (2008), Kafaligonul and Stoner (2010) investigated whether these temporal effects extended to short-range motion as well. The results of their experiments suggested that static sound stimuli modulate the perception of direction and speed of short range motion. Additionally, they showed that there was a possibility that cross-modal temporal interactions may also be occurring as early as in the middle temporal area (MT) (which is a key area in visual processing (Born, Bradley, 2005) and mainly deals with the detection of motion) even though it has been regarded a purely visual area traditionally.

## 2.4 Attention

In every second of our lives, we are constantly bombarded with sensory information from the environment. These sensory inputs are processed with the assistance of prior experiences, current state and sudden presence of information. However, the human brain is incapable of processing all this information due to the high cost of neuronal activity needed for cortical computation and due to limited energy being provided to the brain. Hence, there should be a way to select the crucial information for survival. The key mechanism that allows us to ignore irrelevant information and selectively process key information is called attention or selective attention. The realization of the importance of attention in human life has led to it currently being a significant topic of research in vision and cognitive neuroscience (Carrasco, 2011). Visual attention can be investigated under three main types: spatial attention, feature-based attention and object-based attention.

Spatial attention is a mechanism that allows humans to distribute their attention over a scene (e.g., in the visual field). There are two distinct categories of spatial attention: overt attention and covert attention. Overt attention is allocated by the physical movement of the eye or the head towards the target location in order to put the target object onto the fovea (which is the area in the retina that provides the most acute vision). Covert attention is allocated without any physical movement but by attending to a target on the periphery (Roberts, Summerfield, Hall, 2009). Covert attention also helps us investigate the environment and then direct our eye movement towards relevant content (Carrasco, 2011).

Visuospatial attention is the attentional system that orients visual attention (Posner, 1980) to a particular object or a particular location in space in order to process that information more efficiently (Umiltà, 2000; Vernet et al., 2019). Visuospatial attention can be covertly oriented by using *spatial cues*. In a usual cueing task, participants are asked to detect a signal at one point out of a number of locations. The cue appears before the main stimulus and mostly shows the participant where the stimulus is going to appear. It is possible to create more complicated cueing paradigms by increasing the number of locations and using invalid or neutral (uninformative) cues (Shimozaki, Eckstein, Abbey, 2003). Comparing the performance of participants in the presence of valid, invalid and neutral cues permits the analysis of the effects of cues on attention (Posner, Nissen, Ogden, 1978). There are two main types of cues that shift attention covertly: endogenous cues and exogenous cues. Commonly, endogenous cues are central and symbolic cues that influence attention in a top-down manner. For instance, an endogenous cue can be an arrowhead pointing towards the location that is desired to be cued or it can also be an informative color cue (Mukai et al., 2011; Johnson et al., 2020). On the other hand, exogenous cues are non-symbolic and they work as an automated system for attention modulation. When working with exogenous cues, usually the cue appears suddenly. Hence, it naturally and automatically diverts visuospatial attention to itself (Blurton, Greenlee, Gondan, 2015). Endogenous cues usually need about 300ms to be deployed whereas exogenous cues usually need about 100-120ms to be deployed. Because of the different temporal natures of these cues, endogenous attention is also called *Sustained Attention* since observers can sustain their attention at a specific location as long as the task requires while exogenous attention is called *Transient Attention* since it quickly becomes active and dies (Carrasco, 2011).

It is known and accepted that cues can be used to orient attention. However, there are two different hypotheses with contrasting explanations of what causes the spatial cueing effects. The first hypothesis is the *selective perception hypothesis* which allows the utilization of renewable resources. The selective perception hypothesis assumes that at the cued location the perceptual encoding is increased. Moreover, it is possible to shift the selective processing from one spatial location to another if there is enough time (Denison, Heeger, Carrasco, 2017). The second hypothesis is the *selective decision* hypothesis which assumes that perceptual encoding has unlimited capacity and hence cued and uncued locations are encoded simultaneously. Additionally, cues are used to incrementally weigh the cued information and they are used in making decisions (Johnson, Palmer, Moore, Boynton, 2020). There is still ongoing research to investigate which hypothesis is more valid under different situations and task demands.

### *2.4.1 Filter Models of Attention*

In 1958, Broadbent proposed a model in which from the multitudes of information received, one message is selected according to its physical characteristics and the rest of the information became lost (Broadbent, 1958). His model assumed that attention had a bottleneck design since numerous sensory inputs go in but only the attended information can go out. In the model, all sensory inputs first go through a sensory buffer with unlimited capacity, and then they are filtered based on physical characteristics to prevent the processing system from overloading. Finally, only the inputs with certain physical characteristics are allowed to pass for further cognitive processing. The information that was not selected by the filter remains in the sensory buffer for a short amount of time and then they are discarded if they are not selected by the filter (see Figure 2.7). According to this model, all the semantic processing is done after the filtering process which means that the filter does not take the meaning of the information into account while filtering.

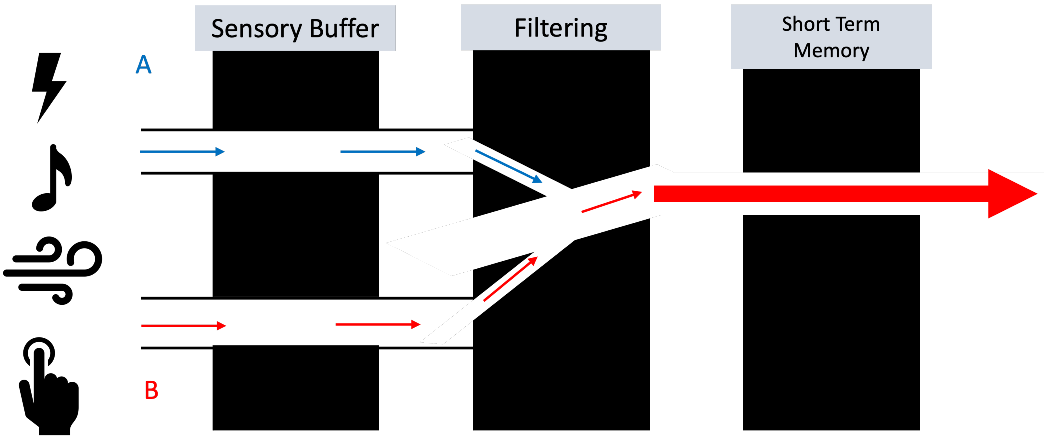


Figure 2.7. Broadbent's filter model of attention. The stimuli first go through a sensory buffer and then, enter a filter where only one of the several stimuli are allowed to pass into the short-term memory. The other stimuli wait for a small amount of time in the sensory buffer and if they are not processed further, they are discarded.

However, Broadbent's model did not explain why some of the unattended stimuli still pass through the filter. To solve this issue, adding to Broadbent's model, Treisman proposed that after the filtering process the information that was not selected by the filter is not discarded but attenuated by the filter for the system to be able to focus on the stronger signal (Treisman, 1964). Figure 2.8 illustrates Treisman's filter model of attention.

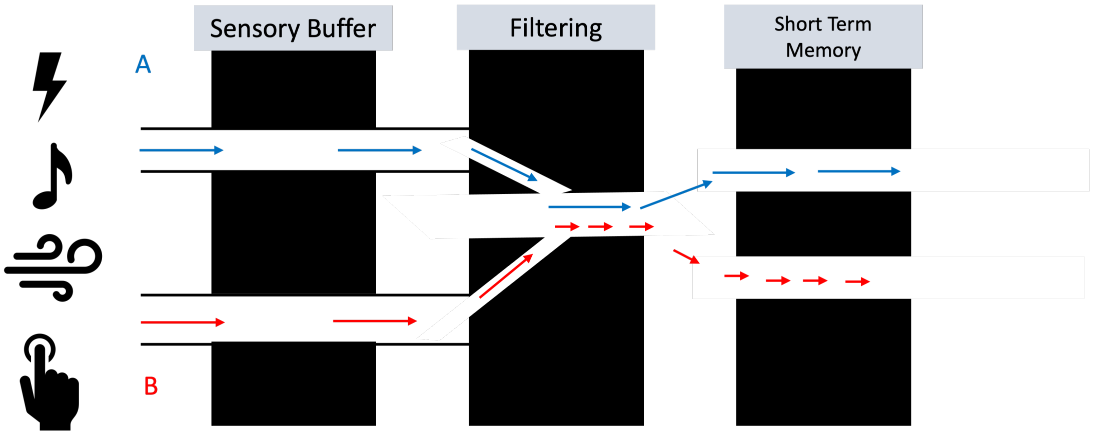


Figure 2.8. Treisman's filter model of attention. Similar to Broadbent's model, the stimuli first go through the sensory buffer and then pass through a filter. However, unlike Broadbent's model, the filter does not stop the task-irrelevant stimulus completely from passing. Instead, it attenuates the task-irrelevant stimulus. Hence, in the further processing steps, the task-relevant stimulus is much more dominant than the task-irrelevant stimulus.

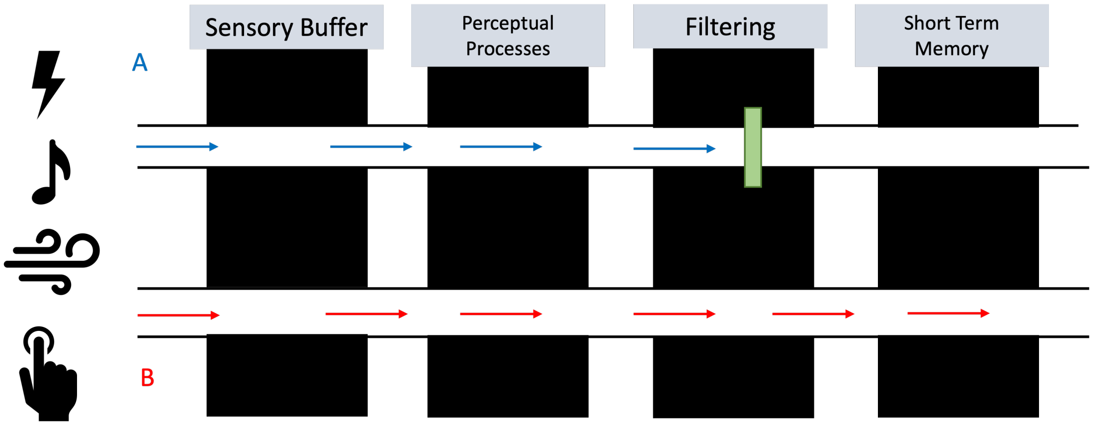


Figure 2.9. Deutsch & Deutsch's filter model of attention. In this model, the order of the processes is different than the previous models. Deutsch & Deutsch propose that the stimuli go through a sensory buffer, but they are processed perceptually before being filtered. The task-relevant stimulus is allowed through the filter and other stimuli are not.

Deutsch & Deutsch proposed a different order of these processes. Unlike Broadbent and Treisman, they proposed that the filtering came after the perceptual and semantic analyses (Deutsch & Deutsch, 1963) which has been termed 'late-filter theory' while the previous theories have been termed 'early filter theory' since the filtering came before any perceptual processes. They stated that all of the received stimuli are processed first and the bottleneck occurs because the individual is not able to respond to all stimuli at once (Cowan, 1995). The filter model proposed by Deutsch & Deutsch is illustrated in Figure 2.9.

### *2.4.2 Limited Resource Model of Attention*

Single resource theory regards all cognitive mechanisms identical in a single resource pool where tasks require a certain amount of resources to be performed. The task performance can be defined specifically by a performance resource function which shows the amount of resources required for a certain task (Pew, Mavor, 1998). If there are two tasks to be performed and there are not enough resources to perfectly perform both of them, then the tasks have to compete for the necessary resources and in the end, one task gets more resources and that task is performed better than the other task.

Multiple resource theory states that there are again limited resources, but these resources are not identical with each other. There are resource pools available for different operations like sensory processing and semantic processing. Each operation acquires the resources they need from the relevant resource pool. Hence, when performing multiple tasks, task performances diminish more when the tasks are similar to each other rather than different since they need to get their resources from the same resource pool (Wickens, 1980). Figure 2.10 shows the properties of single and multiple resource models of attention.

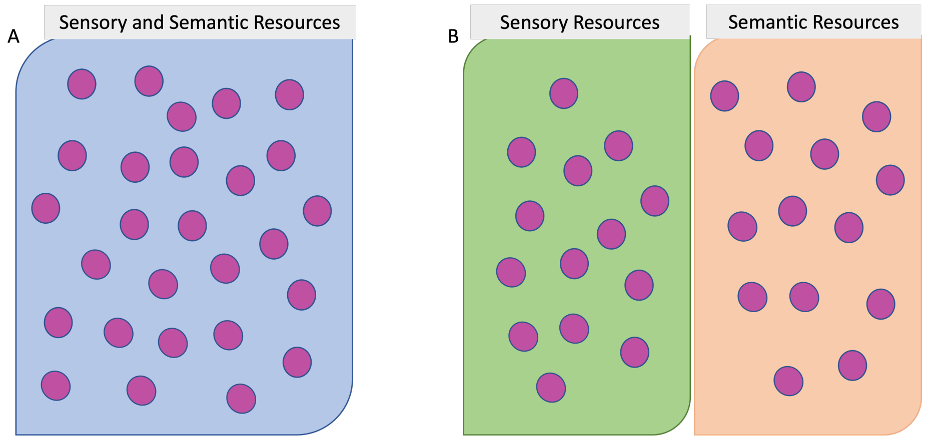


Figure 2.10. (A) Single resource theory. In this theory, the resources for both semantic and sensory processes are in a single pool. (B) Multiple resources theory. In this theory, there are two different pools for sensory and semantic resources. They do not use the resources of each other.

### *2.4.3 Perceptual Load Theory of Attention*

Yantis and Johnson (1990) built on both of these ideas of attention and they suggested that there was a filter that is used to select which stimulus is going to be focused on. However, unlike the previous filter theories, they proposed that the filter can change its location according to the task demands. Perceptual load theory was developed using this idea and its purpose was to determine the factor that changed the location of the filter where this factor was named *the perceptual load* (Murphy, Groeger, Greene, 2016).

Perceptual load theory of attention suggests that both perceptual (external properties) and cognitive (internal properties) loads are effective on selective attention whereas Lavie defined perceptual load as the complexity of physical stimuli (Lavie, Tsal, 1994; Lavie, 1995). Perceptual Load Theory of Attention proposes that the attentional resources are limited. So, the task-relevant stimuli are prioritized and they are processed before task-irrelevant stimuli. Moreover, all of the available resources must be used to the full extent. Hence, if the task-relevant stimulus requires all of the attentional resources, the task-irrelevant stimulus will not be able to be processed since there are no attentional resources left. However, if the task-relevant stimulus requires just a small amount of resources, then the leftover attentional resources will be used for processing the task-irrelevant stimulus (Cartwright-finch, Lavie, 2007).

In this theory both selection mechanisms of early filtering and late filtering are incorporated and which one is used depends on the type of stimulus that is presented. The early selection model is used when distractors or task-irrelevant stimuli cannot be processed because the task has a high perceptual load and the capacity is reached, but the late selection model is used when the perceptual load is low and all present stimuli, whether target or distractor, are processed (Lavie, 2005).

## 2.5 Multisensory Processing and Attention

Since both attention and crossmodal interactions play critical roles in our daily lives in terms of enhancing and organizing the processing of stimuli, it is not extravagant to think that these processes may occur together or influence one another (Donohue et al., 2015). Even though most of the initial studies on attention focused on the visual modality, there is growing interest in the relationship between crossmodal interactions and attention. Besides vision and audition, crossmodal interactions also occur between taste, smell and touch. However, the relation of attention and multisensory integration of these other modalities have not been studied as extensively, and the most commonly studied connection between crossmodal interactions and attention has been in the audiovisual domain (Koelewijn, Bronkhorst, Theeuwes, 2010).

An important line of research on multisensory attention investigated the properties of the temporal window of integration. Donohue et al. (2015) used a stream/bounce paradigm and performed three distinct experiments to see whether attention would narrow or broaden the temporal window of integration during audiovisual processing. The behavioral results suggest that the way attention interacts with multisensory perception is rather flexible. They further showed that attention narrowed the TWI when the task was to report whether the visual stimuli bounced or streamed through each other but it broadened the TWI (see Figure 2.11) when the task was to explicitly judge the simultaneity of the auditory stimulus with the visual stimuli (Donohue, Green, Woldorff, 2015).

In the multisensory attention studies, another specific topic actively examined is to identify modality-specific attentional systems and whether there is a supra-modal attentional system (Blurton et al., 2015). Spence and Driver (1996) investigated the possibility of cross-modal links in endogenous covert orienting in audiovisual interactions. They have found that auditory endogenous attention shifted visual covert orienting (Spence, Charles, Driver, 1996). Another study by Spagna, Wu and Fan (2020) showed that there might be a unified supra-modal entity that aids in the processing of conflicts by using similar mechanisms in both unimodal and cross-modal scenarios.

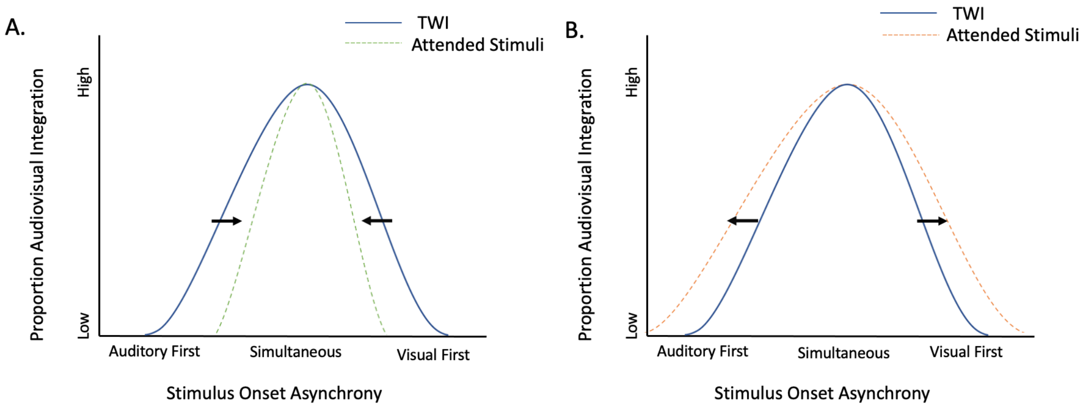


Figure 2.11. Proposed effects of attention on the Temporal Window of Integration. (A) Hypothesis 1 claims that attention narrows the temporal window of integration. (B) Hypothesis 2 claims that attention broadens the temporal window of integration. Adapted from (Donohue, Green and Woldorff, 2015).

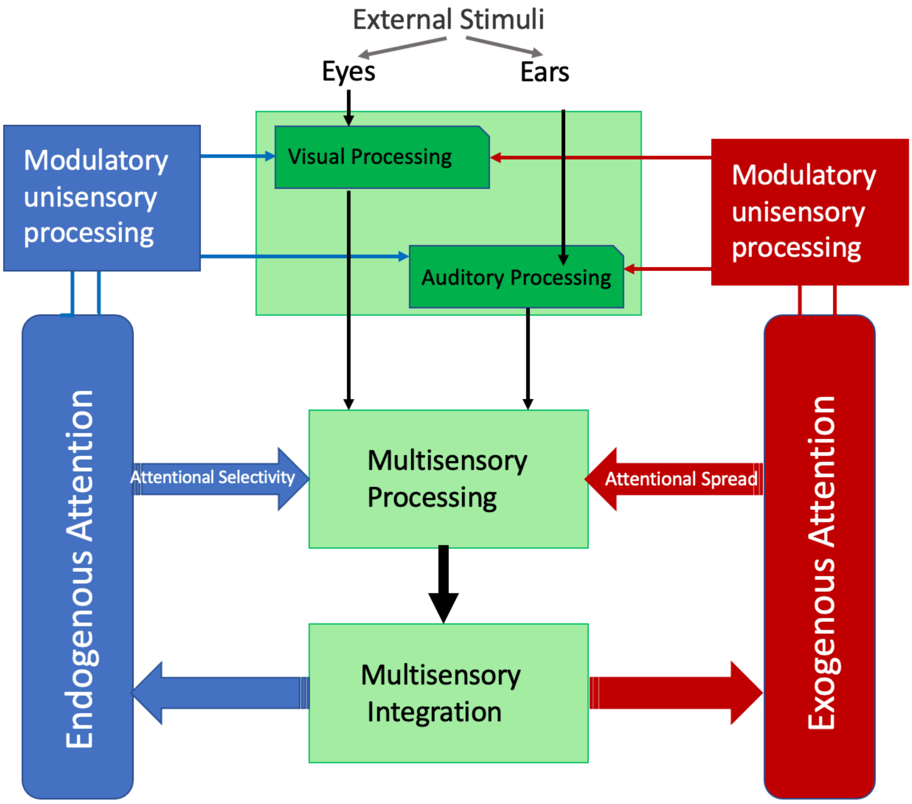


Figure 2.12. Schematic showing the interaction of multisensory mechanism with endogenous and exogenous attention. Adapted from (Tang, Wu, Shen, 2015).

As mentioned in the previous sections both endogenous and exogenous attention, modulate not only unimodal processing but also multisensory processing. Attentional selectivity is the manner that endogenous attention modulates multisensory processes, while exogenous modulates these processes by attentional spread instead. Attentional selectivity determines to what extent the simultaneously presented stimuli can be integrated. Attentional spread happens when there are two stimuli and, endogenous attention can spread from one modality to the other in an exogenous way so that the unattended one becomes the attended one (Tang, Wu, Shen, 2015). Figure 2.12 shows these interactions of multisensory integration with endogenous and exogenous attention.

## 2.6 Bayesian Framework

As we have stated previously, the human brain gets countless stimuli at every second. However, thankfully the perceptual system is capable of integrating these information into a coherent perception of the world, even though it operates by inherently noisy and variable neural signals. The variable nature of the signals make perception a probabilistic process. The perceptual inferences are guided by not only the external stimuli, but also the prior knowledge and experience of the perceiver (Alais, Burr, 2019). This concept of the brain using probabilistic reasoning to integrate information is referred to as the Bayesian Approach, which uses the Bayes' Theorem at its core (Pouget et al., 2013; Bayes, 1783 ;see Equation 2.1 for Bayes' Theorem where P(B|A) is the likelihood and P(A) is the prior). Multisensory integration is one of the operations that best illustrate the probabilistic approach of the brain. We can use Bayesian Framework to model integration of sensory information from different modalities since it is a probabilistic mathematical model that has two components: prior probability (which can represent knowledge that was previously acquired, learning and expectations) and likelihood (which can represent the probability of the stimulus). Hence prior acts as an internal model of statistics of the external environment, while the likelihood acts as the inherently noisy signal (Chen, Vroomen, 2013).

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)} \qquad \text{Equation 2.1}$$

In the integration of a visual and an auditory stimulus, there is one prior (to represent prior knowledge) and there are two likelihoods: one for the visual stimulus and one for the auditory one. So, the probability of an audiovisual integration happening can be calculated by using the Bayes' Theorem as shown in Equation 2.2.

$$P(S_{AV} \mid S_A, S_V) = \frac{P(S_A \mid S_{AV}) P(S_V \mid S_{AV}) * P(S_{AV})}{P(S_A) P(S_V)} \qquad \text{Equation 2.2}$$

Maximum Likelihood Estimation Model (MLE) is a more simplified version of the Bayes' Theorem (it does not take the prior into account) that has been successful in modeling the integration of audiovisual stimuli. MLE model combines the inherently noisy stimuli in an optimal way that, the likelihood of the outcome to reflect the external stimulus accurately is maximized (Alais, Burr, 2019).

Many studies have used Bayesian Framework to investigate and model the integration of auditory and visual stimuli in a spatial context. One of the measures used to examine Bayesian Inference in a spatial context in audiovisual integration is using saccadic eye or head movements. Witten and Knudsen (2005) have reported that psychophysical story results support that the model that the brain weight information according to its reliability. In another study Wanrooji, Bremen and Van Opstal (2010) investigated whether if participants adaptively calculate the expected alignment of audiovisual stimuli during a rapid head-orienting response. Their results showed that the subjects' prior expectation of audiovisual congruency dynamically updated. So, they proposed a model of prior probability estimation to be able to explain their results. Apart from the classical Bayesian or MLE models, more modified models exist as well. For instance in a study, a more specific bayesian model was proposed for audiovisual integration for spatial localization. In this model the MLE model was modified by adding a prior probability distribution which increased the weight of the visual information (Battaglia et. al, 2003).

However there are not as many studies that use the Bayesian Framework to model the integration of auditory and visual stimuli in a temporal context. There is one study that focuses on the temporal aspects of audiovisual integration modeled according to the Bayesian Framework by Hartcher-O'Brien and Alais (2011). In their study they used a version of ventriloquism that was purely temporal to examine whether if temporal mislocalization of audiovisual signals with slight asynchrony would adhere to the MLE model. Their results revealed that even though the MLE model is very successful in modeling multisensory cue combinations in spatial contexts, the accuracy advantages that were predicted by the MLE model does not apply in a purely temporal context.

## 2.7 Current Study

In this thesis, we designed a behavioral study to investigate the relationship between attention and audiovisual processing. More specifically, the experiments were constructed to examine the effect of endogenous cues on the temporal ventriloquism effect using the audiovisual interaction in apparent motion perception. Accordingly, in the experiment participants performed two tasks: a simple speed discrimination task on visual modality and an auditory discrimination task. While investigating the main research question, we also wanted to examine whether task difficulty also affects attention or not in this context. Hence, the experimental design included two different levels (easy vs. hard condition) of auditory task difficulty.

### *2.7.1 Speed Discrimination Task*

A speed discrimination task is based on participants' ability to correctly recognize and distinguish between two or more distinct speeds of motion. Moreover, this task may be utilized in different ways such as discriminating perceived speeds rather than actual speeds of a motion. This perceived speed discrimination paradigm is useful in audiovisual temporal ventriloquism research because it allows researchers to assess whether or not auditory stimuli affect the perceived speed of an apparent motion.

### *2.7.2 Auditory Discrimination Task*

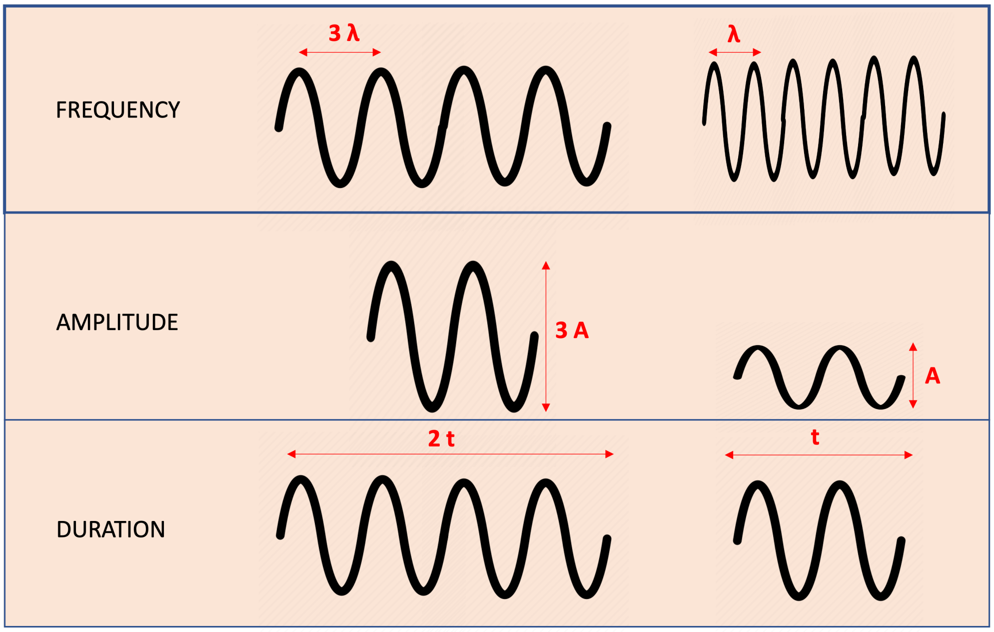


Figure 2.13. A visual interpretation of the main characteristics which can be modified and used in an auditory discrimination task. Frequency is the number of times a sound pressure wave repeats itself per second. Amplitude is the intensity or loudness of a sound. Duration is the length of time that a tone is sounded.

An auditory discrimination task is based on participants' ability to correctly recognize and distinguish between two or more distinct sounds. These tasks can be operationalized over several stimulus characteristics such as frequency, intensity and duration (Jones et. al., 2009). Figure 2.13 depicts the features that can be used in an auditory discrimination task.

### *2.7.3 The Experimental Design*

In this study, participants first completed a visual presentation and some training sessions before starting the main experiment blocks. There were two types of training sessions: visual training and audiovisual training. Audiovisual training consisted of easy and hard training sessions and then they were followed by their respective experiment blocks (easy training was followed by easy experiment blocks and hard training was followed by hard experiment blocks). The order of the experimental blocks was counterbalanced for each participant, such that some participants started their sessions with easy blocks and some participants started with hard blocks. The timeline of the training and experiment blocks can be seen in Figure 2.14.

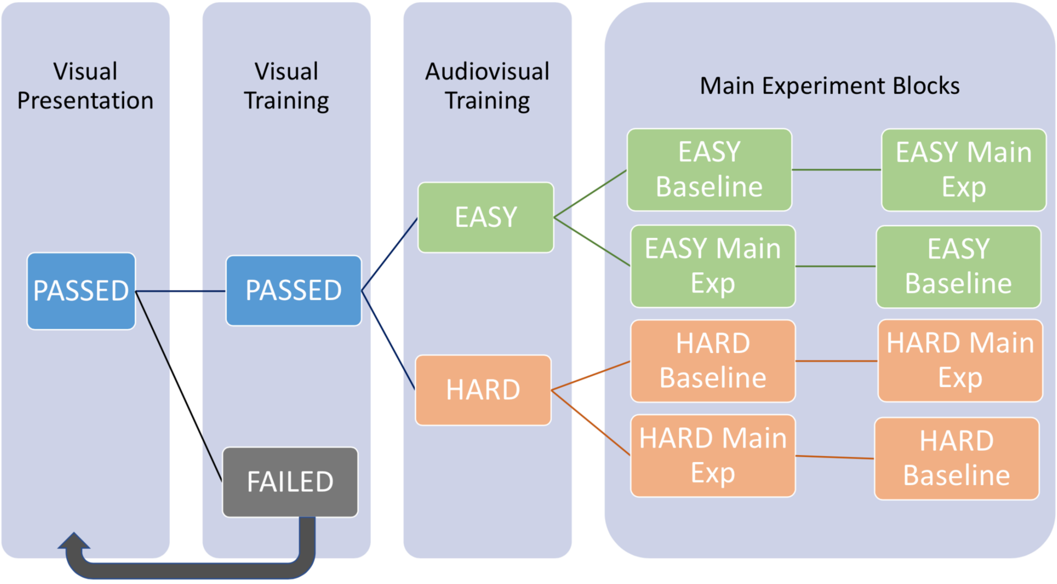


Figure 2.14. Diagram showing the training and experiment blocks. All participants started with a visual presentation of visual stimuli and then continued with the visual training block. In the audiovisual training block, some participants started with an easy training block and then continued with easy experiment blocks, while other participants started with a hard training block and then continued with hard experiment blocks to counterbalance the possible effect of the order of the different experimental blocks might have had on the results.

The training and experiment blocks can be summarized as follows: In the visual presentation block, participants were shown two apparent motions: *fast* and *slow*, and participants were informed which motion was the *fast* motion and which one was *slow* the slow motion. Participants were tasked to remember these motions. In the visual practice block, participants were asked to classify the motions they saw on the screen as *fast* or *slow*. In this block, there were 4 different speeds of motion but the participants were not informed that there were more than two speed conditions in the block. If the participants could classify/categorize the majority of the motion speed correctly, they could continue the experiment. Otherwise, they could not continue. Before the audiovisual training block participants were introduced to two kinds of clicks: one with *low frequency* and another one with *high frequency*.

In the audiovisual training block participants were asked to classify the movement as *fast* or *slow* as they did in the visual practice block. They also were instructed to classify the auditory click they heard as *low frequency* or *high frequency* using specific keys on the keyboard. If the participant classified most of the clicks correctly they were allowed to continue the experiment.

In the baseline block of the experiment, participants performed the same task as in the audiovisual training blocks. In the main experiment block, participants performed the same task as the baseline block again. However, in the main experiment block, there were endogenous cues in the form of red diamonds that appeared either at the right side of the fixation point, left side of the fixation point or at both sides of the fixation point. Participants were informed that these red diamonds (cues) represent where the sound is going to be coming from in the next trial, so if the diamond was on the right side, the clicks were going to be coming from the speaker to the right of the monitor or if the diamond was on the left side of the fixation point the clicks were going to be coming from the speaker to the left of the monitor. If the diamonds appeared on both sides of the fixation point, then the participant was told that the clicks could either come from the speaker on the left or the right of the monitor but they would not know which one until they heard the sound.

In the experiment, we used *temporal ventriloquism effect* and *response time* as measures since in most visual perception studies response time for the trials are used to measure the efficiency of attention (Murphy, Groeger, Greene, 2016). How these measures were calculated is explained in detail in chapter 3.3.

The hypothesis of both of the experiments was that endogenous cues would influence the audiovisual integration of audiovisual stimuli. Specifically, it was hypothesized that valid cues would facilitate the audiovisual interactions and binding. Hence, they would increase the temporal ventriloquism (i.e., auditory time interval) effects on perceived visual speed. Moreover, participants would take more time to complete the task for neutral cues since they are not informative like valid cues. Lastly, it was predicted that the auditory task difficulty would increase the temporal ventriloquism effects since it would force the participant to focus more on the auditory task and also increase the response times because a harder task typically requires more time to complete. Similar to the temporal ventriloquism effect, it was hypothesized that the response times would be longer for the neutral condition than the valid cues since valid cues provided information about the spatial location to attend. Finally, the hypothesis stated that the response times would be longer for the harder auditory task.

# CHAPTER 3

## 3 METHOD

This chapter is dedicated to the experiment of my thesis work that had the main goal of investigating the effect attention might have on the interaction of auditory and visual processes in the brain. As mentioned in chapter 2, crossmodal interactions are one of the most important functions of the human brain since they enhance the processing of external stimuli via different modalities. In addition, attention is another essential mechanism of sensory processing that filters all stimuli from the external environment and allows us to focus on the relevant content and stimulation. Hence, this experiment was designed to examine the interaction of these two crucial mechanisms of the perceptual system. The hypothesis proposed that the temporal ventriloquism effect would be higher in the valid cue trials than the neutral cue trials since valid cues inform the participant of the location to attend in the spatial domain. Moreover, it was hypothesized that the response times in neutral cue trials would be longer than valid cue trials. The reason for this assumption was the fact that neutral cues did not provide any spatial information while valid cues provided spatial information about the auditory clicks. The participants were assumed to be able to discriminate between the two clicks more easily (hence, more quickly) when the cue called attention to the auditory clicks. Additionally, it was hypothesized that the temporal ventriloquism effect would be higher in the hard experiment blocks than the easy experiment blocks. This was assumed because increased task difficulty would compel the participant to attend more to the task at hand. Moreover, it was assumed that the response times would be longer in the hard experiment blocks than in the easy experiment blocks. This was expected since participants can take more time to respond to a harder task than an easy task.

### 3.1 Participants and Apparatus

Fourteen observers (7 females; mean age of 25.36 +-3.06 *SD* years, 22-34 years range) participated in the experiment. All observers had normal or corrected-to-normal visual acuity and normal hearing. None of them had a history of neurological disorders. Participants gave informed consent, and all procedures were in accordance with international standards (Declaration of Helsinki, 1964) and approved by the ethics committee at Bilkent University. One participant’s data were not included in the statistical analysis because of ceiling in the observed temporal ventriloquism effect.

Matlab version 2015 (MathWorks, Natick, MA) and Psychtoolbox 3.0 were used for stimulus presentation and data acquisition (Brainard, 1997; Pelli, 1997). To present the visual stimuli, a 20-in CRT monitor with 1280x1024 pixel resolution and 100Hz refresh rate (Mitsubishi Diamond Pro 2070SB ) was used at a 57cm viewing distance. The luminance calibration and gamma correction of the display were made via a SpectroCAL photometer (Cambridge Research Systems, Rochester, Kent, UK). Auditory stimuli were presented via 2-channel speakers positioned next to the monitor and were 90 cm apart with the midpoint at the fixation on the screen. The sound pressure levels were set to 63 dB and were measured with a sound-level meter (SL-4010 Lutron, Lutron Electronics, Taipei, TW). A digital oscilloscope (Rigol DS 10204B, GmbH, Puchheim, Germany) was used to verify if the timing of the auditory and visual stimuli were correct. The experiment room was dark and silent during all of the experiments.

## 3.2 Stimuli and Procedure

The background used during all blocks of the experiment was black (0.561 cd/m$^2$). The fixation point was a white diamond (side length = 0.5°, 108 cd/m$^2$) and the participants were asked to fixate during all of the experimental blocks. In the experiments, a bar (horizontal side length = 0.6°,vertical side length = 1.87°, 56.8 cd/m$^2$) was presented above and diagonal to the fixation point. The distance from fixation to the center of the bar was 3.85°. An apparent motion was achieved by either presenting the bar first at the left position and then the right position (left-to-right apparent motion), or presenting the bar first at the right position and then the left position (right-to-left apparent motion). Each bar was shown for 50ms. To have a coherent apparent motion percept with different speed conditions, we used time intervals ranging from 20 to 180 ms between the two bars. To make it easy to understand and remember, the apparent motions with different speed conditions were named *slow, medium-slow, medium, medium-fast* and *fast* apparent motions. Table 1. lists the basic features of different apparent motion types.

In addition to the visual stimuli, auditory clicks were introduced during each presentation of each apparent motion. Each click had a 20ms duration and was only introduced through one speaker, located either on the left or right of the monitor. The location of the auditory click was randomized for each trial. In the easy blocks, each low-frequency click was constructed by a rectangular 480 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization, and was introduced at 63dB sound pressure level (SPL). Similarly, each high-frequency click was constructed by a rectangular 1 kHz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization and was introduced at 63dB sound pressure level (SPL). Table 2 lists the properties of the auditory stimuli in easy blocks.

Table 1. The list of visual time intervals to have apparent motions with different speed conditions.

| APPARENT MOTION NAME | TIME INTERVAL BETWEEN TWO BARS (ISI) |
|---|---|
| **SLOW** | 180 ms |
| **MEDIUM-SLOW** | 140 ms |
| **MEDIUM** | 100 ms |
| **MEDIUM-FAST** | 60 ms |
| **FAST** | 20 ms |

Table 2. The names of different auditory stimuli, their respective frequencies and the time interval between two clicks in a single trial of an easy block.

| AUDITORY STIMULUS NAME (EASY) | FREQUENCY | TIME INTERVAL BETWEEN TWO BEEPS (ISI) |
|---|---|---|
| **LOW FREQUENCY INNER SOUND** | 480 Hz | 100 ms |
| **LOW FREQUENCY OUTER SOUND** | 480 Hz | 160 ms |
| **HIGH FREQUENCY INNER SOUND** | 1 kHz | 100 ms |
| **HIGH FREQUENCY OUTER SOUND** | 1 kHz | 160 ms |

Compared to the easy blocks, the difference between the frequencies of the auditory stimuli was significantly less in the hard blocks. While designing these blocks, various amounts of differences between the *low* and *high* frequencies were investigated. However since the sensitivity of audition in every individual is different, it was possible that participants would not be able to easily perceive a difference less than 200 Hz, because the auditory stimuli were very short and played on a single speaker. Hence, 640 Hz and 840 Hz were selected to be the frequencies of the *low* and *high* frequency sounds in the hard blocks (see Table 3). In the hard blocks, each low-frequency click was constructed by a rectangular 640 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization and was introduced at 63dB sound pressure level (SPL). Finally, each high-frequency click was constructed by a rectangular 840 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization and was introduced at 63dB sound pressure level (SPL).

In each trial, two consecutive clicks were concurrently played with each presentation of apparent motion. However, in a trial, the interstimulus interval (auditory time interval) between the first click and the second one could either be 100 ms (*inner sound condition*) or 160 ms (*outer sound condition*).

Table 3. The names of different auditory stimuli, their respective frequencies and the time interval between two clicks in a single trial of a hard block.

| AUDITORY STIMULUS NAME (HARD) | FREQUENCY | TIME INTERVAL BETWEEN TWO BEEPS (ISI) |
|---|---|---|
| **LOW FREQUENCY INNER SOUND** | 640 Hz | 100 ms |
| **LOW FREQUENCY OUTER SOUND** | 640 Hz | 160 ms |
| **HIGH FREQUENCY INNER SOUND** | 840 Hz | 100 ms |
| **HIGH FREQUENCY OUTER SOUND** | 840 Hz | 160 ms |

There were also endogenous cues that indicated the location of the auditory stimuli during a trial. The cues were red diamonds (side length = 1.06°, 35000 cd/m$^2$) that were horizontally 4° away from the fixation point. The cues were used in only the main experimental sessions. In each trial of these sessions, the onset timing between the cue and the first apparent motion was 450 ms and the cue duration was 100 ms. In this block, we used *valid cues* that called attention to the speaker that was going to play the auditory click. This was accomplished by displaying a red diamond next to the fixation point on the same side as the active speaker in that trial. Additionally, we used *neutral cues* including red diamonds on both sides of the fixation point. In this way, in a neutral trial, the participant was not given any spatial information about the sound location. Instead of invalid cues we decided use neutral cues because we did not want to modulate the temporal ventriloquism effect negatively, but rather we wanted to see if cues strengthened the integration or not. All visual stimuli are shown in Figure 3.1.

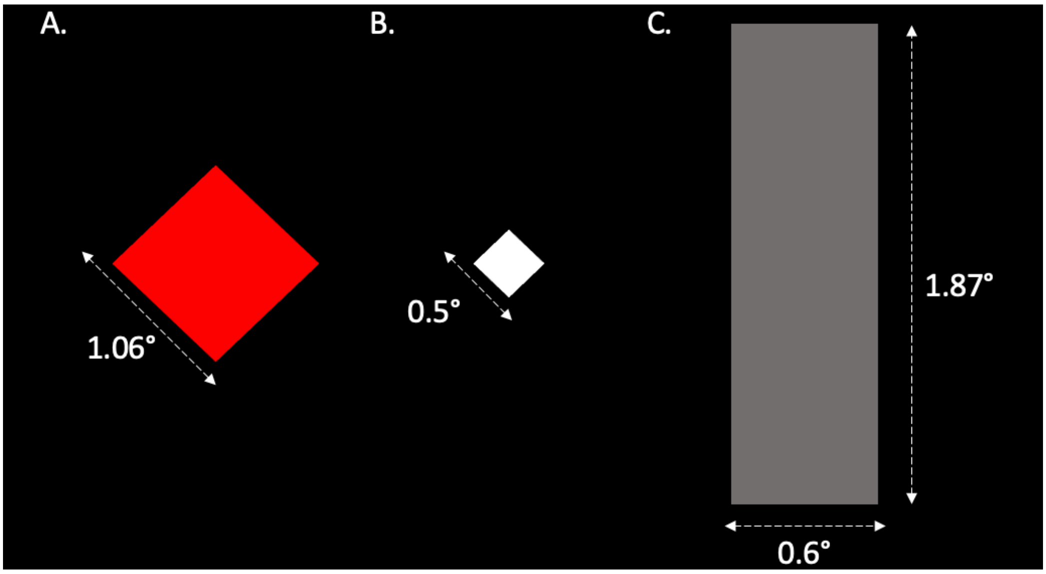


Figure 3.1 (A) the cue (red), (B) the fixation point (white) and (C) the bar (gray) used as the visual stimulus.

The experiment was divided into 4 main blocks: visual presentation, visual training, audiovisual training and main experiment. There were two types of audiovisual training (*easy* and *hard*) and four types of main experiment blocks (*easy* baseline, *hard* baseline, *easy* experiment, *hard* experiment). Before starting the experiment participants were shown the fixation point and were instructed to always fixate on it.

### *3.2.1 Visual Presentation Block*

In the visual presentation block, participants were shown only the *fast* and *slow* apparent motions. In a single visual presentation block, there were 30 trials in which the order of *fast* and *slow* trials was randomized. In each trial, the participants were informed about two extreme categories of movement speed (20 ms vs. 180 ms). For instance, when the apparent motion with movement speed of 20 ms was presented at a trial, the participant was informed that this was the *fast* motion. Similarly, when the apparent motion with movement speed of 180 ms was presented, the participant was informed that this was the *slow* motion. The participant was asked to observe the trial and learn the kind of apparent motion by listening to the feedback given by the experimenter. After the participant stated that they have grasped the two types of motion, they were asked to name what kind of motion was presented in each trial. The visual presentation block lasted 30 trials. Figure 3.2 shows the timing diagram of the visual presentation block

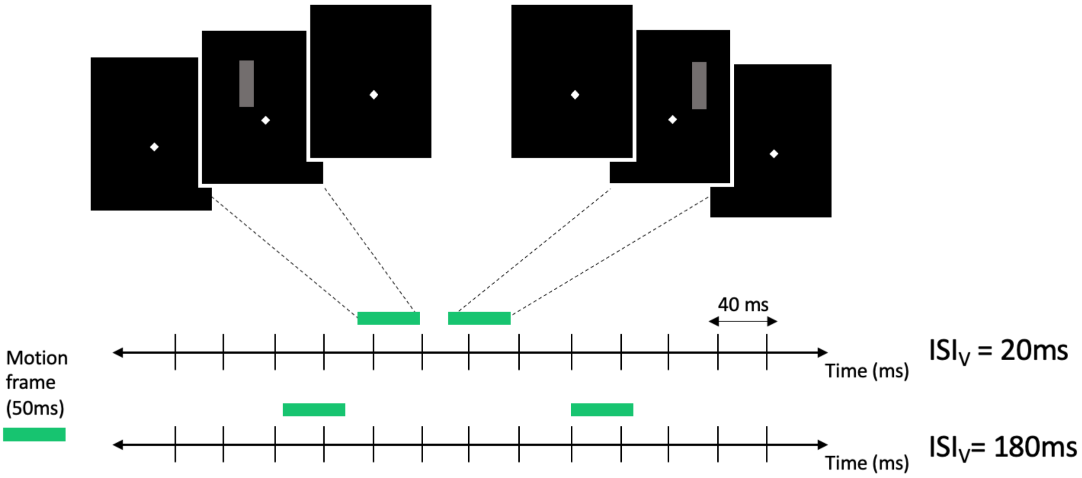


Figure 3.2 The timing diagram of the visual presentation block. In this block, there were two different speed conditions (fast:20 ms ISI, slow: 180 ms ISI) of apparent motion that was shown to the participant. They were trained to identify/classify these types of motion.

### *3.2.2 Visual Training Block*

After the visual presentation block, each participant completed a visual training block. The main purpose of having this block was to evaluate whether a participant could correctly classify/categorize apparent motion speeds according to the instructions. This block included 4 different speed conditions for the apparent motion: *fast, medium-fast, medium-slow and slow* (Table 3) with the time intervals shown in Table 4, ordered from *fast* to *slow*. The participant was not informed that there were 4 different speed conditions, they were only instructed to classify the motions in each trial as *fast* or *slow*. We applied the following criteria: the passing condition for this session was that the participant classified at least 50% of *medium-fast* & at least 75% of *fast* trials as fast, and at least 50% of *medium-slow* & at least 75% of *slow* trials as slow. Participants who failed to satisfy these criteria were not included in the further steps of the study. Each training block had 32 trials.

In this block, participants were given feedback by the color change of the fixation point. The color of the fixation point changed to green (116000 $\text{cd/m}^2$) at the end of each trial if the participant classified the movement speed correctly. If the participant failed to classify the trial correctly, the fixation point remained white. Participants were informed that they had 2 seconds in each trial to respond via the keyboard. The ‘T’ and ‘Y’ buttons were used to respond to the question of the task. Participants were asked to press ‘T’ if they thought the movement was *fast* and press ‘Y’ if they thought the movement was *slow*. Participants were also provided a picture of a keyboard layout with the 'T, Y, G, H' buttons were marked with their meanings in the experiment. The experimenter made sure that the participants understood the button combination by doing mock trials until the participant stated that they were comfortable with the button layout. The visual training session had 30 trials. Figure 3.3 shows the timing diagram of the visual practice block.

Table 4. Time interval information used to have different apparent motion speeds in the visual training block

| **Apparent Motion Time Interval** | **Auditory Click Time Interval** |
|---|---|
| 20 ms | N/A |
| 60 ms | N/A |
| 140 ms | N/A |
| 180 ms | N/A |

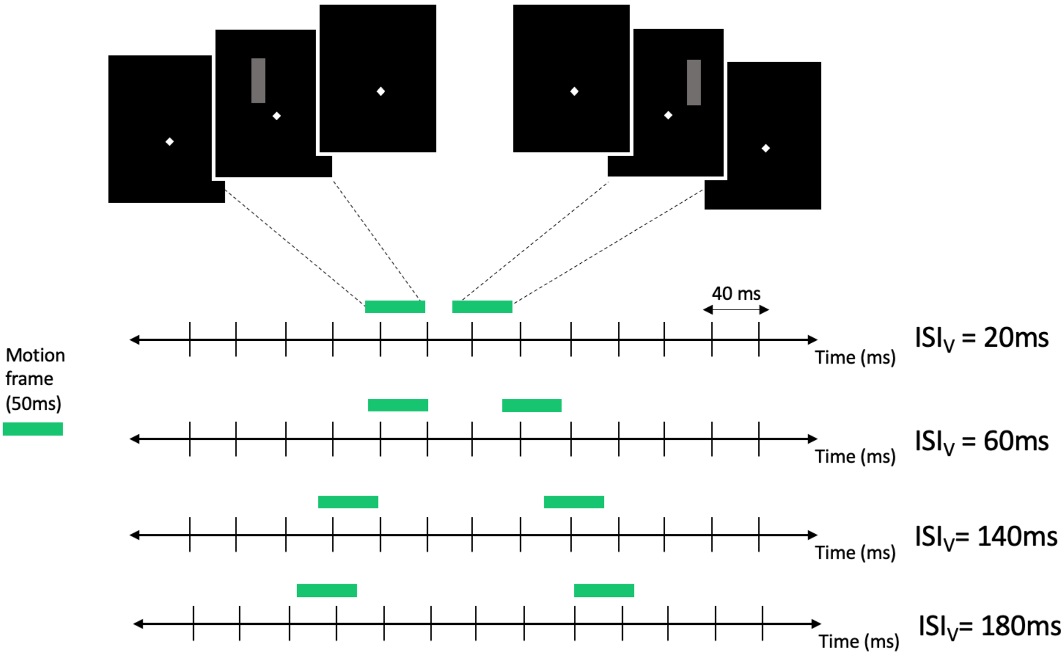


Figure 3.3 The timing diagram of the visual practice block. In this block there were four different speeds of apparent motion. The *fast* motion had 20 ms in between the two motion frames that put a rectangle bar on the screen. The *medium-fast* motion had a 60 ms gap, the *medium-slow* motion had a 140 ms gap and lastly, the *slow* motion had a 180 ms gap in between the two motion frames.

### *3.2.3 Easy Audiovisual Training Block*

The main goal of this block was to assess whether a participant could correctly classify the *low frequency* and *high frequency* clicks during each trial, but to train the participants to use the input keys correctly, we also tasked them with the speed categorization task as well. As in the visual training block, participants were given feedback through the fixation point color which turned green if the answer given in the trial was correct and remained white if the answer was wrong or the response time exceeded 2 seconds. The trials included an apparent motion and a simultaneous click sound. All trials had apparent motions with the same ISI, which was 100 ms, so as not to create a temporal ventriloquism effect and interfere with the actual experiment. This block also contained some catch trials in which the ISI of the apparent motions could be 20 ms or 180 ms. The click sounds that accompanied these apparent motions all had the same ISI with their respective apparent motions. The ISI information for all the apparent motions and auditory clicks in the training blocks can be seen in Table 5. Figure 3.4 shows the timing diagram of the easy audiovisual training blocks.

Table 5. Time interval information of the apparent motions and the auditory clicks that accompany them in the easy and hard audiovisual training blocks

| Apparent Motion Time Interval | Auditory Click Time Interval |
|---|---|
| 20 ms (catch trial) | 20 ms |
| 100 ms | 100 ms |
| 180 ms (catch trial) | 180 ms |

All participants were introduced to the *low* and *high* frequency clicks before starting the audiovisual training block. The clicks were played in random order and participants were asked to categorize each click (two-alternative forced-choice, *low* vs. *high*). Using a few trials, participants discriminated sounds to check whether they were able to perform according to the instructions. After this initial step, the participants were informed that they had two seconds in each trial to respond by using the keyboard. The ’T’, ’Y’, ‘G’ and ‘H’ buttons were used as response keys. As shown in Figure 3.5, each keypress corresponded to each of the four different options. When participants were instructed during this phase, these combinations were introduced based on location on the keyboard. The ‘upper’ buttons T and Y were used to take input for *high frequency* clicks and the ‘lower’ buttons G and H were used to take input for *low frequency* clicks to make it easier for the participants

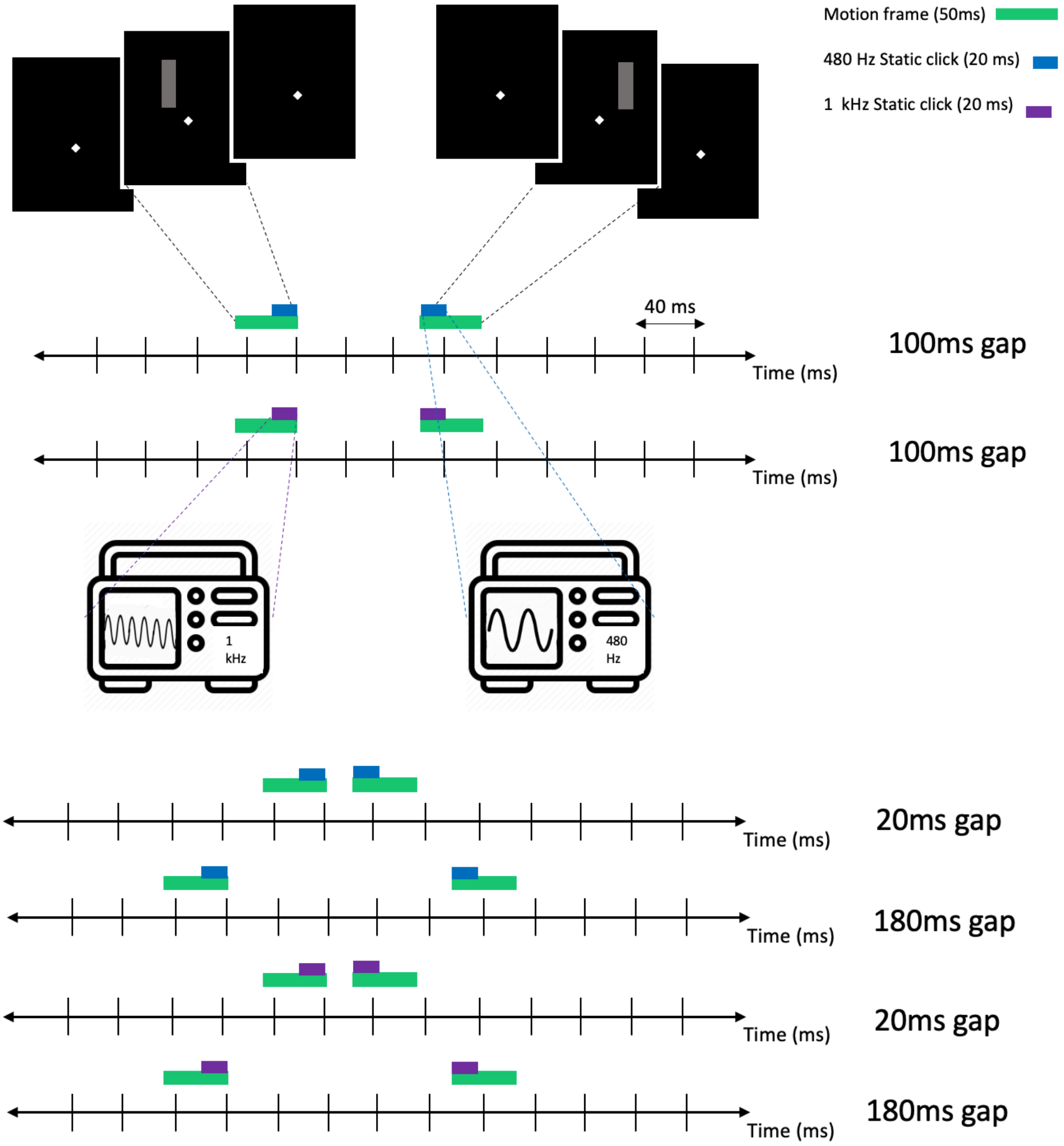


Figure 3.4 Timing Diagram of the easy audiovisual training block. The time interval between the two motion frames in regular trials was 100 ms while the time intervals between the two motion frames in catch trials could either be 20 ms (for fast motion) or 180 ms (for slow motion). All possible regular trials and catch trials are shown in the figure. In this block a 480 Hz static click was used as the low frequency click and a 1 kHz static click was used as the high frequency click.

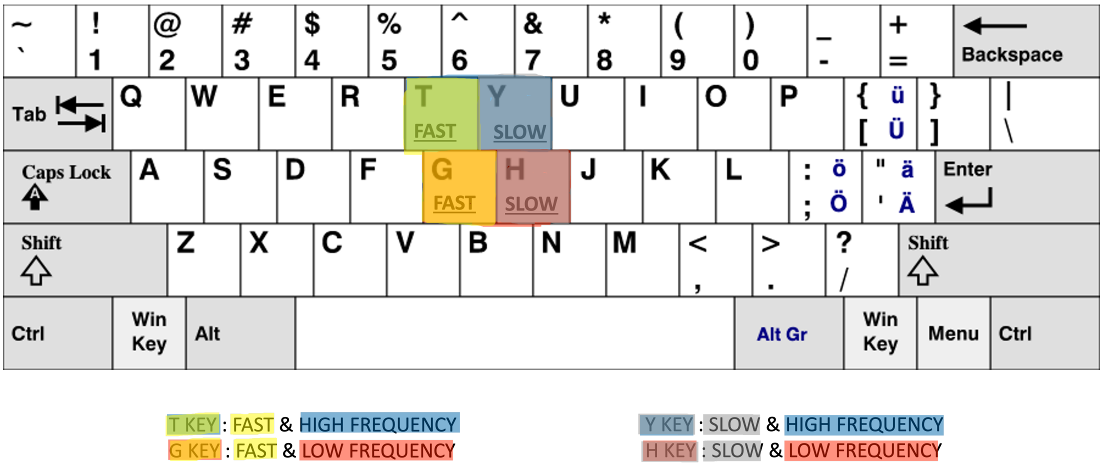


Figure 3.5 In this figure, the input button layout is shown on a keyboard. The buttons T, Y, G and H were used in the audiovisual training and the experiment blocks. The ‘upper’ buttons T and Y were used to take input for *high frequency* clicks and the ‘lower’ buttons G and H were used to take input for *low frequency* clicks to make it easier for the participants to remember the button combination. For the same purpose, the buttons ‘on the left’ T and G were used to take input for *fast* motion and the buttons ‘on the right’ Y and H were used to take input for *slow* motion.

to remember these button combinations. The buttons ‘on the left’ T and G were used to take input for *fast* motion and the buttons ‘on the right’ Y and H were used to take input for *slow* motion. Participants were also provided with a picture of a keyboard layout with the ’T, Y, G, H’ buttons were marked with their meanings. The experimenter made sure that the participants understood the button combination by doing mock trials until the participant stated that they were comfortable with the button layout. We applied the following criteria: the passing condition for this session was that the participant classified at least 50% of *medium-fast* & at least 75% of *fast* trials as fast, and at least 50% of *medium-slow* & at least 75% of *slow* trials as slow. Participants who failed to satisfy these criteria were not included in the further steps of the study. The audiovisual training session included 30 trials. After this session, the results of each participant were screened to check whether he/she satisfied the criteria.

### *3.2.4 Hard Audiovisual Training Block*

In accordance with the easy audiovisual training block, the main goal of this block was to check if the participant could correctly classify the *low frequency* and *high frequency* clicks in the trials, so the tasks were the same. As in the visual training session, participants were given feedback through the fixation point color which turned green if their answer was correct and remained white if their answer was wrong or the response time exceeded 2 seconds. The trials included an apparent motion and simultaneous click sounds with frequencies shown in Table 3.

The apparent motion types, catch trials and auditory clicks that accompanied these apparent motion types all had the same properties as the ones in the easy blocks. The difference between the easy audiovisual training block and this audiovisual training block was the frequencies of the sounds. In this block, as the *low frequency* sound we used a click with 640 Hz frequency and as the *high frequency* sound we used a click with 840 Hz frequency. Details of how these sounds were generated were described at the beginning of this chapter. Figure 3.6 shows the timing diagram of the hard audiovisual training blocks, while the ISI information for all the apparent motions and auditory clicks in these blocks can be seen in Table 5.

### *3.2.5 Easy Main Experiment Session*

Main Experiment Sessions consisted of two sub-blocks: Baseline and Experiment. The order of these blocks was randomized for each participant and hence some participants started with the baseline block and other participants started with the experiment block to counterbalance any possible confounding factor due to the block order.

In the Baseline block, the trials only consisted of the fixation point and visual apparent motion. In this block, the time interval between apparent motion frames was 100 ms corresponding to the medium level of speed, but the participants were not informed about the speed movement. There were two kinds of sounds: a click with 480Hz frequency called the *low frequency* sound and a click with 1kHz frequency called the *high frequency* sound. Two different time intervals were used to generate the two clicks that were presented in a single trial. *Inner Sounds* had 100 ms between the two clicks and *Outer Sounds* had 160 ms between the two clicks in a single trial. The auditory time intervals are shown in Table 6 and they were used both for the baseline blocks and the main experiment blocks. Unlike the training blocks, the baseline and main experiment blocks did not have catch trials. Baseline blocks had 64 trials, while experiment blocks had 240 trials each. Figure 3.6 shows the timing diagram of the easy baseline blocks.

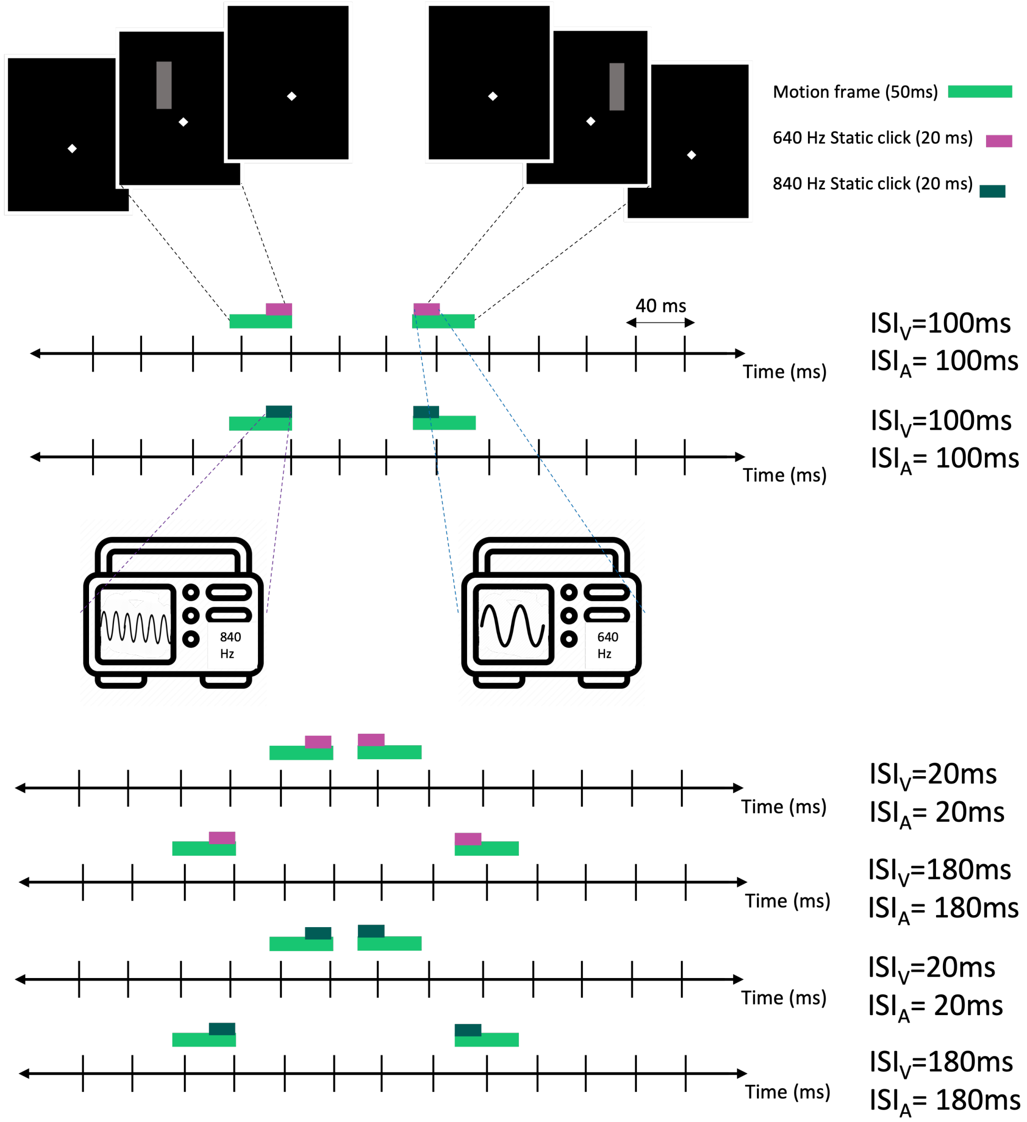


Figure 3.6 Timing diagram of the hard audiovisual training block. The time interval between the two motion frames in regular trials was 100 ms while the time intervals between the two motion frames in catch trials could either be 20 ms (for *fast* motion) or 180 ms (for *slow* motion). All possible regular trials and catch trials are shown in the figure. In this block a 640 Hz static click was used as the *low frequency* click and an 840 Hz static click was used as the *high frequency* click.

Table 6. Interstimulus Interval information of the visual and auditory stimuli in the experiment. Apparent motion only had one ISI while clicks were either 100ms or 160 ms apart from each other in a single trial.

| Apparent Motion ISI | Auditory Beep ISI |
|---|---|
| 100 ms | 100 ms (Inner Sound) |
| 100 ms | 160 ms (Outer Sound) |

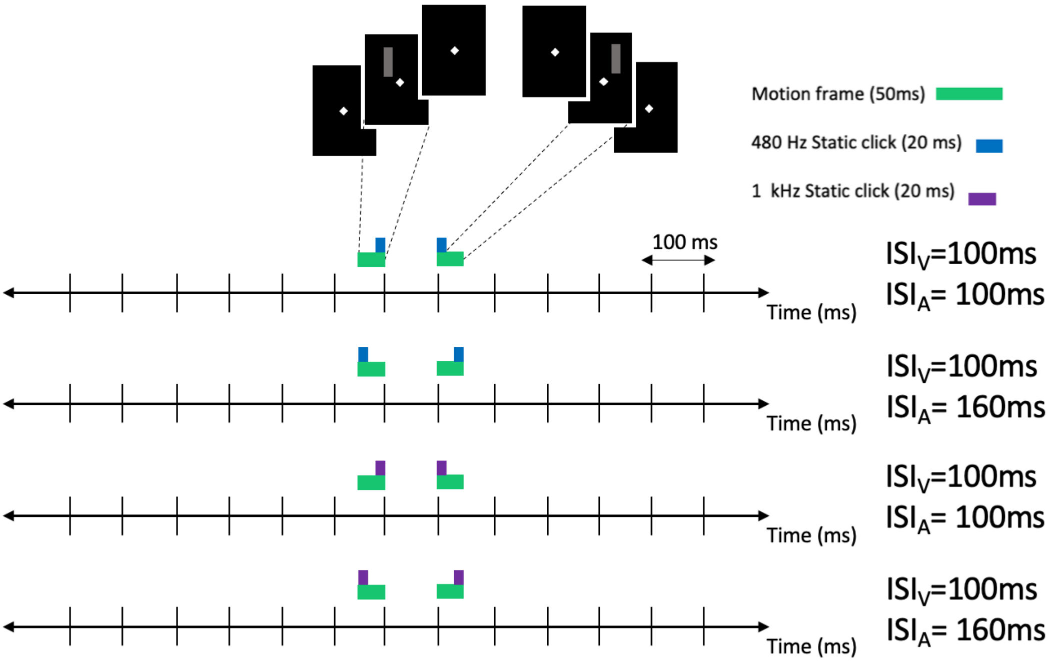


Figure 3.7 The timing diagrams of the baseline block of the easy experiment. In this block there were inner sounds (time interval between two clicks = 100 ms) and outer sounds (time interval between two clicks = 160 ms) that were presented to the participant simultaneous to the apparent motions. The auditory clicks could either be low frequency (480 Hz) or high frequency (1 kHz). The time interval between the two motion frames in the apparent motion was always 100 ms.

Before starting the main experiment session, the participants were also briefly informed about the basic stimulation profile. The cues were introduced to all participants before the experiment. Participants were informed that if the cue showed up to the right of the fixation point then the sound would be coming from the speaker to the right side of the monitor (*valid cue*, right) and similarly if the cue showed up to the left of the fixation point then the sound would be coming from the speaker to the left side of the monitor (*valid cue*, left). All participants were also told that sometimes cues showed up on both the right and left side of the fixation point simultaneously, which meant that the sound could come from either the speaker on the right side of the monitor or on the left side of the monitor but we did not have information which one it will be exactly (*neutral cue*). The timing diagram of the main experiment is depicted in Figure 3.7, where the possible cue conditions and the timings of the *inner* and *outer sounds* are shown.

### *3.2.6 Hard Main Experiment Session*

As described previously the main experiment sessions consisted of two sub-blocks: Baseline and Experiment. The order of these blocks was randomized for each participant and hence some participants started with the baseline block and other participants started with the experiment block to counterbalance any possible confounding factor due to the block order.

As it was done in the easy baseline blocks, in the hard baseline block, the trials only consisted of the fixation point and visual apparent motion. In this block, the time interval between apparent motion frames was 100 ms corresponding to the medium level of speed, but the participants were not informed about the speed movement. There were two kinds of sounds: a click with 640Hz frequency called the *low frequency* sound and a click with 840Hz frequency called the *high frequency* sound. Two different time intervals could be between the two clicks in a single trial. *Inner Sounds* had 100 ms between the two clicks and *Outer Sounds* had 160 ms between the two clicks in a single trial. The auditory time intervals that are shown in Table 6 were used both for the hard baseline blocks and the hard main experiment blocks. The hard baseline and main experiment blocks did not have catch trials either. Figure 3.8 shows the timing diagram of the easy baseline blocks.

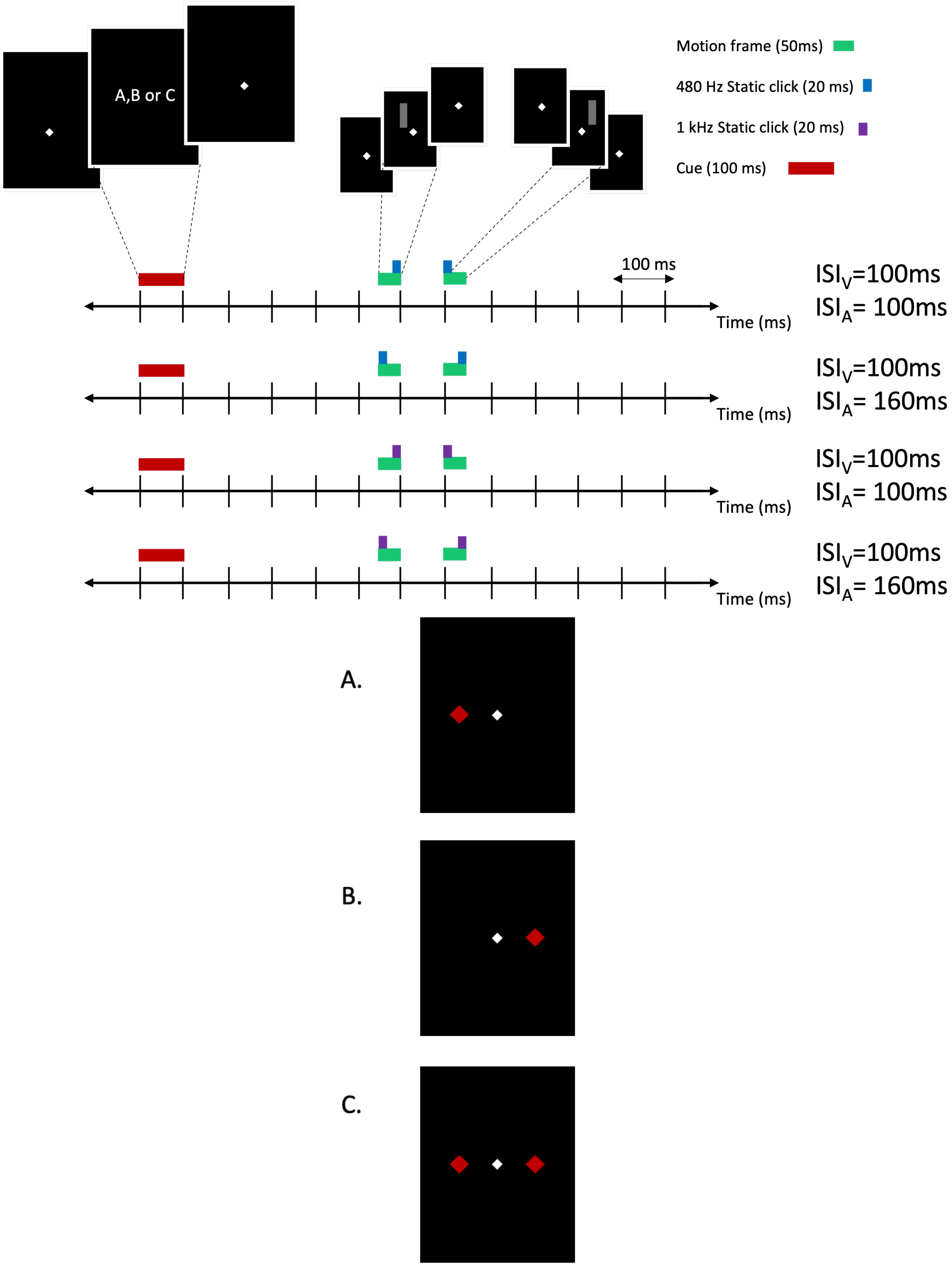


Figure 3.8 Timing diagrams of the easy block of the main experiment. The cue was a red diamond that could appear in the positions shown in (A), (B) or (C). The cue always appeared 450ms before the apparent motion. (A) *Valid Cue* on the left of the fixation point. (B) *Valid Cue* on the right of the fixation point. (C) *Neutral Cue* on both sides of the fixation point.

Before starting the main experiment session, the participants were also briefly informed about the basic stimulation profile again. Participants were informed that if the cue showed up on the right of the fixation point then the sound would be coming from the speaker on the right side of the monitor (*valid cue*, right) and similarly if the cue showed up on the left of the fixation point then the sound would be coming from the speaker on the left side of the monitor (*valid cue*, left). All participants were also told that sometimes cues showed up on both the right and left side of the fixation point simultaneously, which meant that the sound could come from either the speaker on the right side of the monitor or on the left side of the monitor but we did not have information which one it will be exactly (*neutral cue*). The timing diagram of the main experiment is depicted in Figure 3.9, where the possible cue conditions and the timings of the *inner* and *outer sounds* are shown.

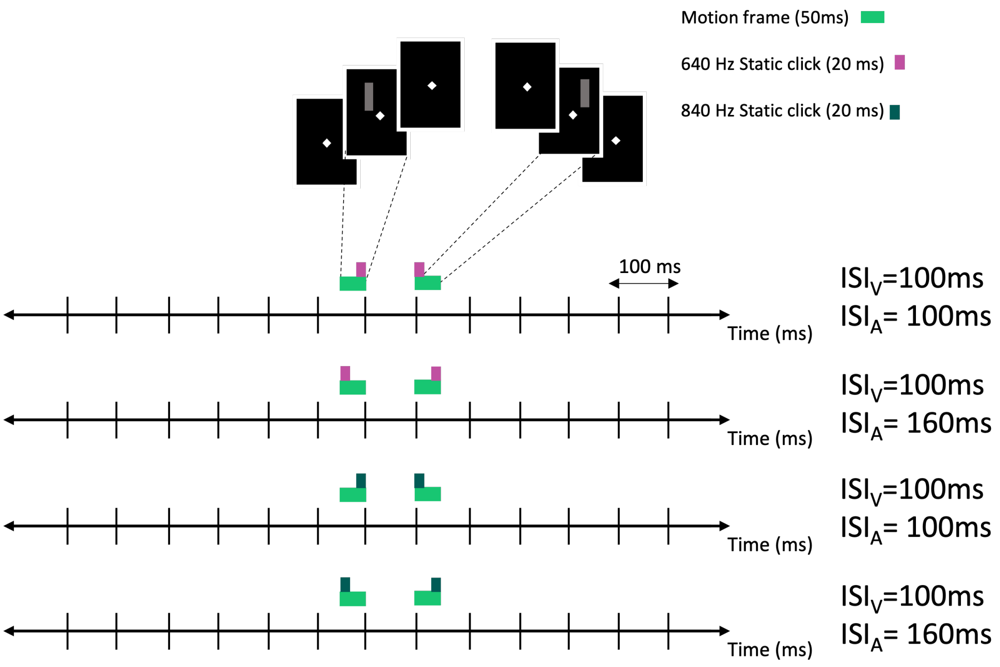


Figure 3.9 The timing diagrams of the baseline block of the hard experiment. In this block there were inner sounds (time interval between two clicks = 100 ms) and outer sounds (time interval between two clicks = 160 ms) that were presented to the participant simultaneous to the apparent motions. The auditory clicks could either be low frequency (640 Hz) or high frequency (840 Hz). The time interval between the two motion frames in the apparent motion was always 100 ms.

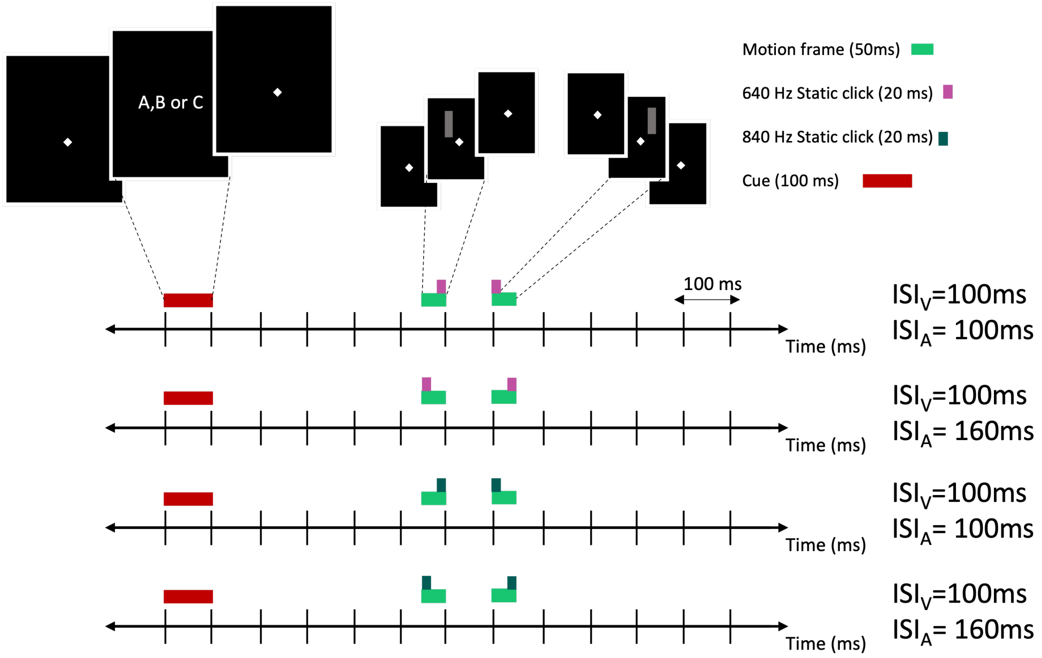


Figure 3.10 Timing diagrams of the hard block of the main experiment. The cue was a red diamond that could appear in the positions shown in A, B or C. The cue always appeared 450ms before the apparent motion. A, B and C were the same as those in the easy block of the main experiment, hence they can be seen in Figure 3.7.

## 3.3 Behavioral Data Analysis

Each different block of participant data was recorded in separate text files. Data was collected and recorded using MATLAB and Psychtoolbox 3.0. Data analysis was performed via using MATLAB and IBM SPSS Statistics (IBM, 2019) scripts that I wrote.

During each block, both responses and response times of each trial were recorded. *Percentage of Fast Speed Responses* (PFR) was calculated by dividing the number of trials that participants responded as FAST by the total number of trials and then multiplying this proportion with 100. Equation 3.1 shows the calculation method.

This PFR value was used to quantify the *Temporal Ventriloquism* effect for each auditory time interval condition.

$$PFR = (FAST_{participantresponse\#}/Trial\_Count) * 100 \qquad \text{(Equation 3.1)}$$

### *3.3.1 Temporal Ventriloquism Effect*

The temporal ventriloquism effect was calculated using *Percentage of Fast Speed Responses (PFR)* of *inner sounds* and *outer sounds*, calculated separately for valid and neutral trials. To calculate the temporal ventriloquism effect in the valid trials, the percentage of *outer sounds* in valid trials that were classified as *fast* ($PFR_{valid}$) was subtracted from the percentage of *inner sounds* in valid trials that were classified as *fast* and the calculation can be seen in equation 3.2.

$$TV_{valid} = PFR_{valid,inner} - PFR_{valid,outer} \qquad \text{(Equation 3.2)}$$

To calculate the temporal ventriloquism effect in the neutral trials, the percentage of *outer sounds* in neutral trials that were classified as *fast* ($PFR_{neutral}$) was subtracted from the percentage of *inner sounds* in neutral trials that were classified as *fast* and the calculation can be seen in equation 3.3.

$$TV_{neutral} = PFR_{neutral,inner} - PFR_{neutral,outer} \qquad \text{(Equation 3.3)}$$

Additionally, the PFR of the baseline sessions were also recorded to be able to compare the cases of valid and neutral trials with the baseline trials which only had visual stimuli with *medium* apparent motion speed and no auditory stimuli.

### *3.3.2 Response Times*

Throughout all the blocks, the response latency of each participant was also recorded. In all trials, response times were based on the onset of visual stimulation and response registration. Participants had a 2 second time limit to answer in each trial. If the participant failed to answer within this period, that trial was discarded and added to the end of the upcoming trial list. Hence, all response times were below 2 seconds in the final dataset. The valid time interval available to the participants is shown in Figure 3.11.

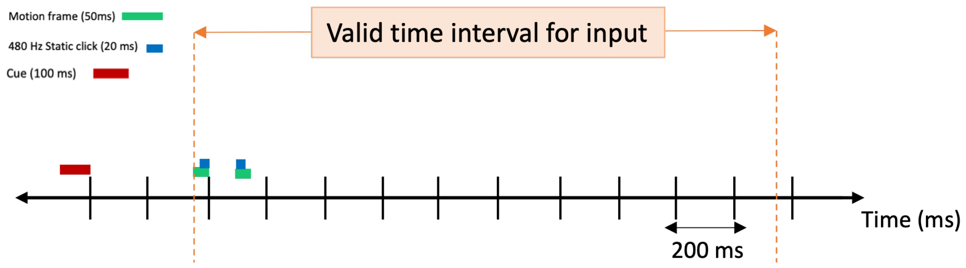


Figure 3.11 The valid time interval of input during a trial. Participants had 2 seconds to input their response using the valid keys ’T, Y, G, H’ mentioned in Figure 3.5.

# CHAPTER 4

## 4 RESULTS

The experiment aimed to investigate the influences of endogenous cues on the temporal ventriloquism effect which was achieved through static clicks during the presentation of visual apparent motions. That is to say, this experimental design was focused on the effects of endogenous attention cues on the perceived speed of an apparent motion which was manipulated by static clicks with different time intervals. Additionally, blocks with different auditory categorization task difficulties were used to see whether the auditory task difficulty had an impact on the temporal ventriloquism effect.

In the following subsections, the behavioral results of the visual practice block and the easy and hard audiovisual practice blocks, the temporal ventriloquism effect, response times, and associated statistical outcomes are provided.

### 4.1. Visual Training Results

The purpose of this block was to evaluate whether a participant could correctly classify/categorize apparent motion speeds according to the instructions. We hypothesized that the percentages of classifying the trial as '*fast*' of all of the trials (*fast, medium-fast, medium-slow and slow motions*) would decrease gradually since as we go from fast to slow, the motions look less alike. The results revealed a gradual decrease in the '*fast*' classification as the ISI in the apparent motion increased (i.e. the apparent motion speed decreased) just as we expected. The mean of all participants also satisfied the passing criteria of this block.

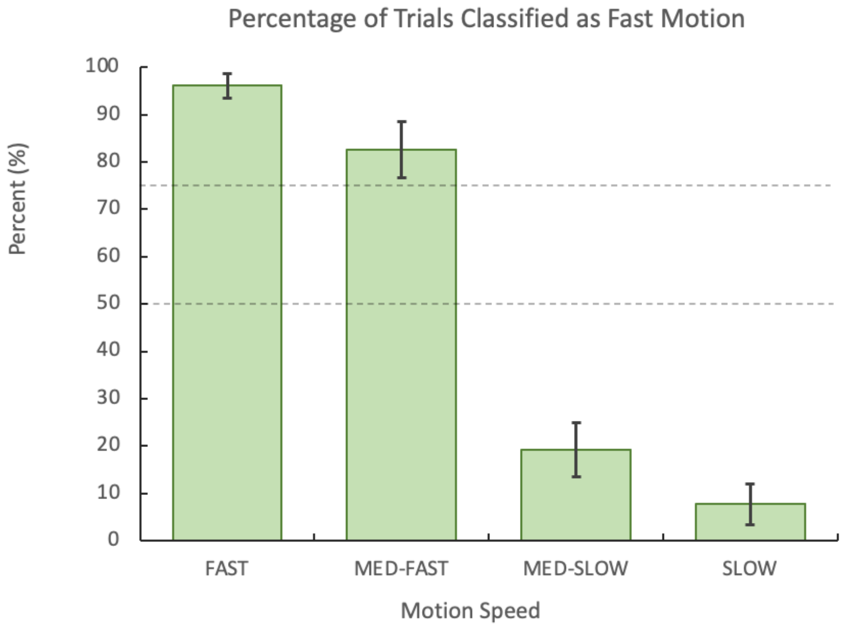
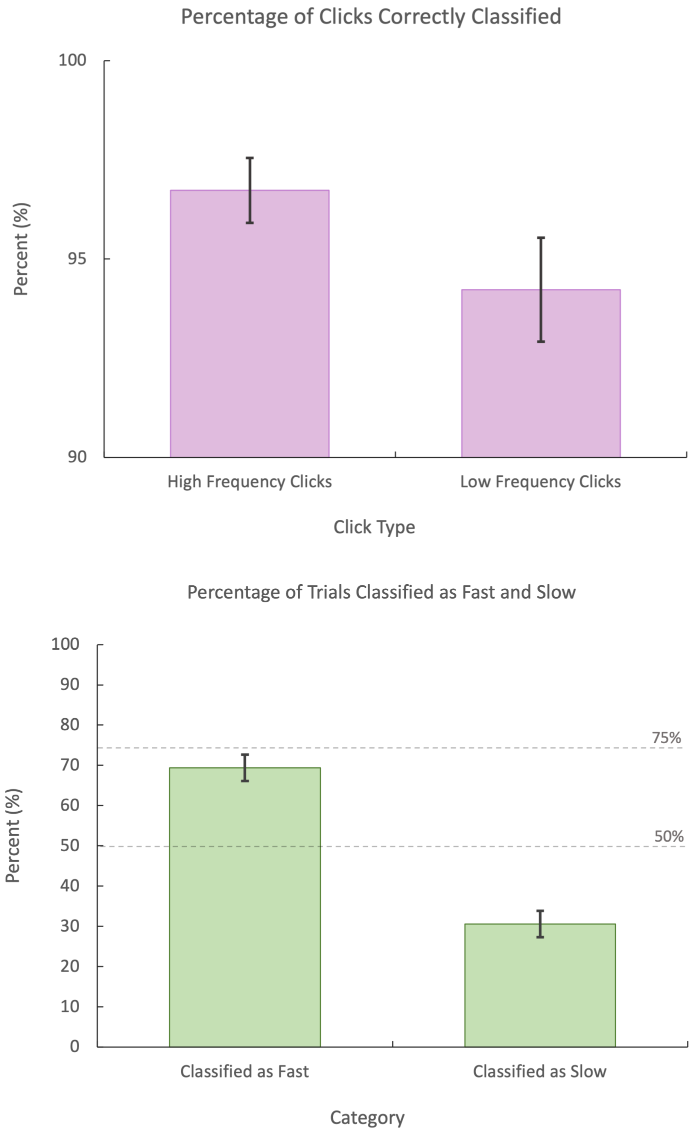


Figure 4.1. Visual Training Results. All participants satisfied the passing criteria for the visual training block. Additionally, the results revealed a gradual decrease in the '*fast'* classification as the ISI in the apparent motion increased (the motion speed decreased).

## 4.2 Audiovisual Training Results

The main goal of the easy and hard audiovisual training blocks was to assess whether participants could correctly classify the *low frequency* and *high frequency* clicks during each trial. In the easy audiovisual training blocks, the results revealed over 90% correct classification of the auditory stimuli type. The mean of all participants satisfied the passing condition of this block. Participants classified the apparent motion as *fast* 72% of the trials while they classified the motion as slow only 28 % of the trials (see Figure 4.2). Meanwhile, in the hard audiovisual training, the results revealed over 90% correct classification for high frequency clicks while the *low frequencies* were correctly detected less than 90% of the trials. The mean of all participants satisfied the passing condition of this block as well. Participants classified the apparent motion as *fast* 63% of the trials while they classified the motion as *slow* only 37% of the trials (see Figure 4.3).



Percentage of Clicks Correctly Classified

Percentage of Trials Classified as Fast and Slow

Figure 4.2 In this figure the Easy Audiovisual Training results are shown. All participants satisfied the passing criteria for this block. The click types were almost 100% correctly classified.

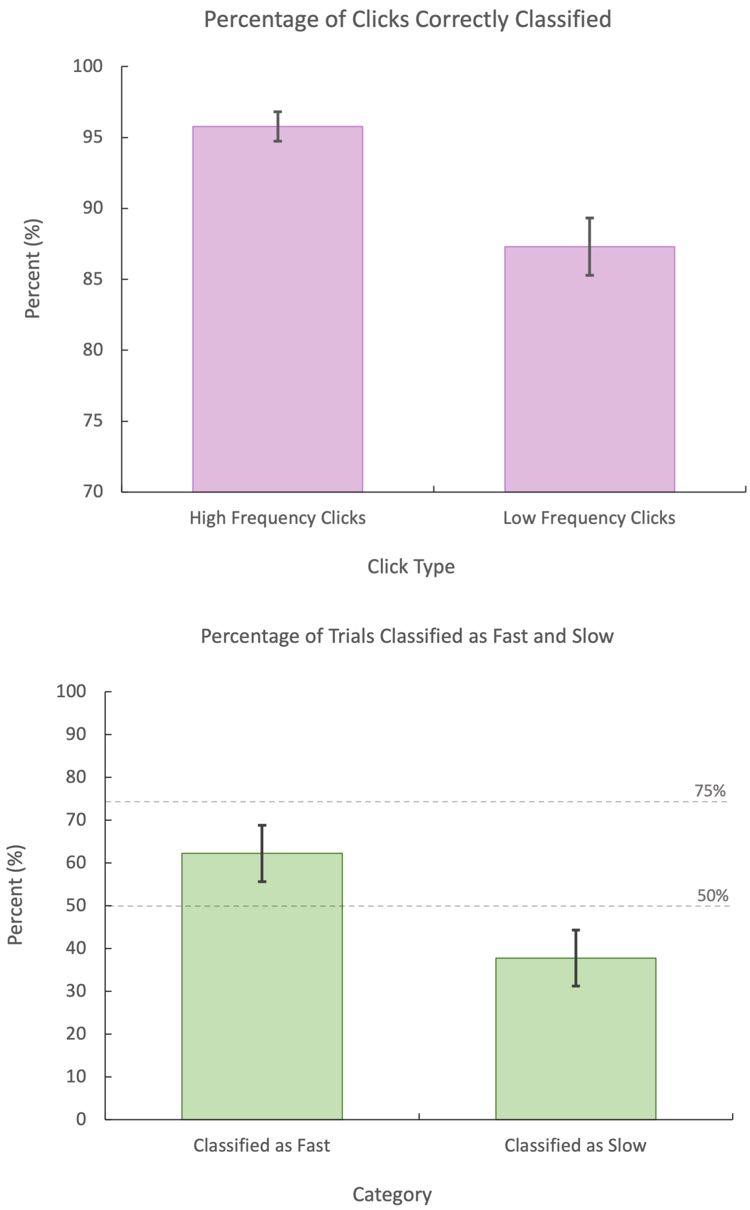


Figure 4.3 In this figure the Hard Audiovisual Training results are shown. All participants satisfied the passing criteria for this block. The high frequency clicks were almost 100% correctly classified while the correct classification percentage of the low frequency clicks decreased compared to the easy audiovisual training blocks.

## 4.3 Temporal Ventriloquism Effect

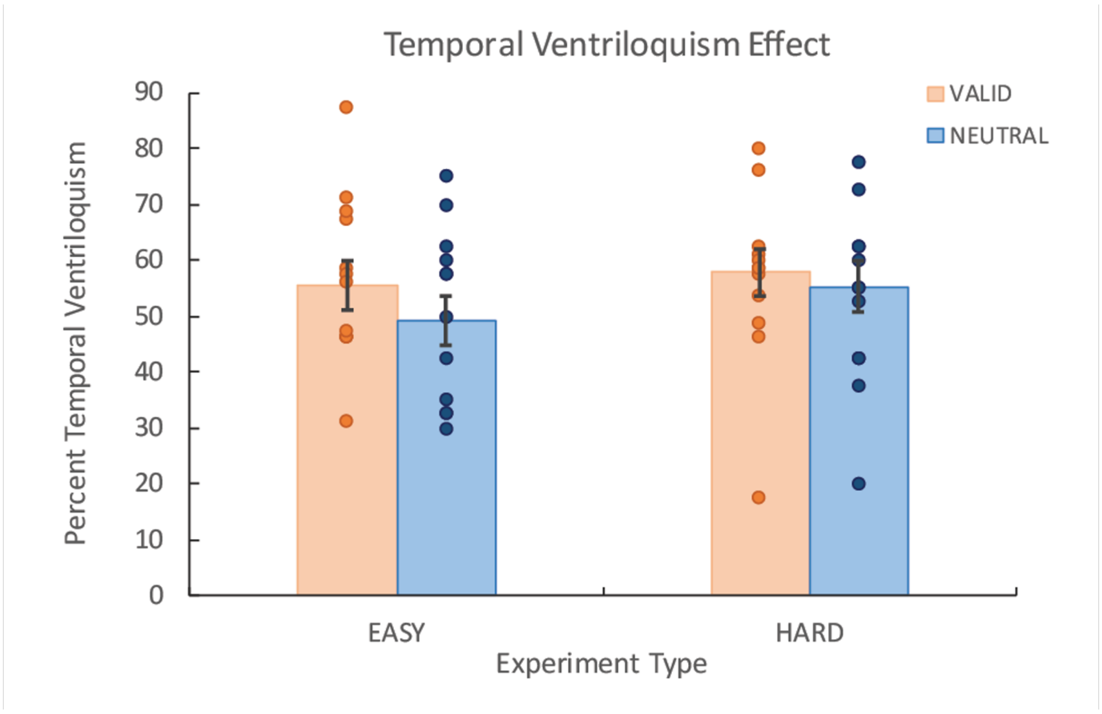
To see whether cueing and difficulty had an effect on the temporal ventriloquism effect, we performed a two-way repeated-measures analysis of variance (ANOVA) with difficulty and cue-type as factors on the averaged percent temporal ventriloquism values. The ANOVA revealed a significant main effect of difficulty on temporal ventriloquism effect ($F_{1,12} = 7.280$, $p = .019$, $\eta_p^2 = .378$), while there was no main effect of cue or the interaction between the two factors (cue: $F_{1,12} = .765$, $p = .399$, $\eta_p^2 = .060$, $\text{interaction}_{\text{cue\&difficulty}}$: $F_{1,12} = .815$, $p = .384$, $\eta_p^2 = .064$). This result displays that the difficulty manipulation we have used in the experiment was effective and it influenced the responses of the participants. Regardless of the cue condition, the temporal ventriloquism effect was higher in the hard trials than in the easy trials. When we visually inspected the results (see Figure 4.4), we saw that there could be a significant effect of the cue on the temporal ventriloquism effect in the easy block, so we performed t-tests to see if there was an effect of the cue.

A paired samples t-test with cue as a factor was performed on the temporal ventriloquism values for the blocks with the easy auditory categorization task. The results revealed a significant effect of cueing on data ( $t_{12} = 2.212$, $p = .047$, Cohen's d = .613), indicating that when the auditory discrimination task was easy, the valid cue enhanced these audiovisual interactions in the temporal domain, and hence suggesting facilitation for audiovisual binding in time. Similarly, a paired samples t-test with cue as a factor was performed on the temporal ventriloquism values for the blocks with the hard auditory categorization task. However, the test results did not reveal any effect of cue on the temporal ventriloquism effect ($t_{12}$= 1.130 , p = .281, Cohen's d = .313)

Table 7. The percentage values corresponding to temporal ventriloquism effect on perceived speed.

| | **Temporal Ventriloquism Effect (%)** | | |
|---|---|---|---|
| **Experiment Blocks** | **Valid Cue** | **Neutral Cue** | **Difference Valid - Neutral** |
| **Hard** | 57.78 | 55.19 | 2.59 |
| **Easy** | 55.38 | 49.23 | 6.15 |
| **Difference Hard - Easy** | 2.4 | 5.96 | |

Figure 4.4 The percentage values of temporal ventriloquism shown for all the cue and task



difficulty conditions (n = 13). Dots represent the data from individual participants. Error bars correspond to ±SEM.

The behavioral results indicated that the percentage values associated with the temporal ventriloquism effect of perceived speed were higher for valid endogenous cues than neutral cues, although only in the easy task on the sound discrimination there was a significant main effect of the cue condition. In terms of task difficulty, the two-way repeated-measures ANOVA test revealed a main effect of difficulty for temporal ventriloquism effect where the effect was higher in the hard difficulty than the easy difficulty as expected in the hypothesis.

As shown in Table 4.1, the difference between the hard experiment block and the easy experiment block was more distinct in the neutral cue case. This might be because the increase in task difficulty forces the participant to attend more to the stimuli, hence increasing the audiovisual processing. However, in the hard experiment block, the *percent temporal ventriloquism effect* difference between *valid cues* and *neutral cues* is less compared to the difference in the easy experiment block.

## 4.4 Response Times

As we did for the temporal ventriloquism effect, we performed a two-way repeated measures analysis of Variance (ANOVA) with difficulty and cue-type as factors was performed on the averaged values. ANOVA did not reveal any main effect of cue or difficulty or any interaction effect (cue: $F_{1,12}$ =1.606, $p = .229$ , $\eta_p{}^2 = .118$, difficulty: $F_{1,12} = 3.028$, $p = .100$ , $\eta_p{}^2 = .201$, interaction of cue and difficulty: $F_{1,12} = .841$, $p = .377$, $\eta_p{}^2 = .066$).

Table 4.2 displays the response time results. As expected by our original hypothesis, response times were longer for *neutral cues* than *valid cues* even though statistical tests did not reveal any significant main effect of the cue in either of the blocks with different auditory task difficulties (see Figure 4.5). Additionally, as the hypothesis predicted, the response times were longer in the hard blocks, the experiment with the harder task, than in the easy blocks. However, again statistical tests did not reveal a significant effect of difficulty.

Table 8. Response Times (ms) Comparison Table

| | **Response Time (ms)** | | |
|---|---|---|---|
| **Experiment Blocks** | **Neutral Cue** | **Valid Cue** | **Difference Neutral - Valid** |
| **Hard** | 942.21 | 935.88 | 6.23 |
| **Easy** | 906.68 | 882.54 | 24.13 |
| **Difference Hard - Easy** | 35.53 | 53.43 | |

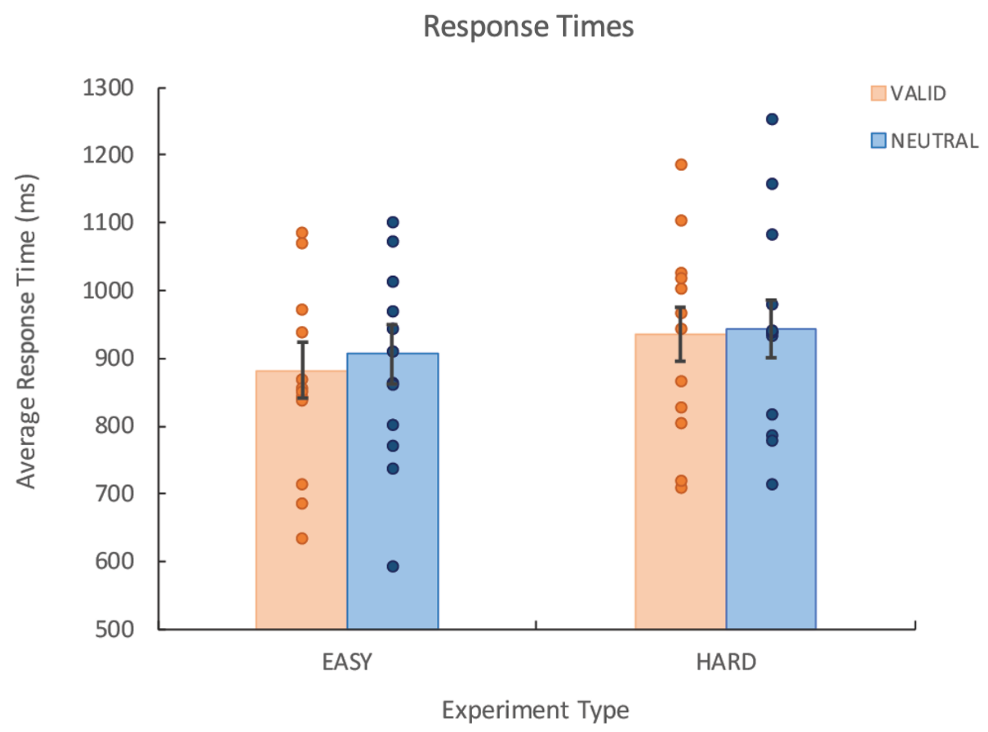


Figure 4.5 Mean response times of valid and neutral cues for all auditory task difficulty conditions (n = 13). The mean response time was longer for neutral endogenous cues than valid cues. Dots represent the data from individual participants. Error bars correspond to ±SEM.

# CHAPTER 5

## 5 DISCUSSION AND FUTURE DIRECTIONS

Sensory perception is critical in human life since it provides information from the outside world, warns against threats and danger, and helps us interact/communicate with the external environment. Given that information is provided via multiple sensory organs, it is important to understand the principles underlying interactions across senses and the contribution of these interactions to the final percept. Every time we talk to someone we are not only listening to the sounds they make, but we watch their lip and body movements as well. I believe the pandemic made more people realize these audiovisual interactions since people had to wear masks that cover their noses and mouths. Research done on the effects of masks on communication during the Covid-19 pandemic also states that the ability to see the mouth of the person speaking is important for people to comprehend what is being told to them (Campagne, 2021; see also Mheidly et al., 2020). Other instances of audiovisual interactions were also mentioned in chapter 2 where either auditory information influenced visual processes or vice versa. The importance of senses is clear, but we are constantly surrounded by stimuli from all modalities. It is unlikely that all of the information is processed at the same time due to limited resources in the cortex. Attention is a critical mechanism that allows us to select and focus on the relevant information from the outside world via our senses.

The main goal of the present study was to investigate whether endogenous attention affects audiovisual processing via spatially informative endogenous cues in the auditory domain. Accordingly, we used an apparent motion paradigm that elicits temporal ventriloquism effects on perceived speed (i.e., auditory time interval) and presented endogenous cues before simultaneous visual apparent motions and concurrent static clicks.

We hypothesized that the endogenous cues influence these audiovisual interactions in the temporal domain. More specifically, we expected that the valid cues indicating the location of the static clicks facilitate auditory signals and hence auditory time interval effects on perceived speed. We also anticipated that it would take more time (i.e. longer response time) to complete the perceptual task under the neutral cue condition since no information was provided about the location of auditory stimuli in these conditions. Moreover, the observed effects of cueing may change based on the task difficulty.

The behavioral results indicated that the percentage values associated with the temporal ventriloquism effect of perceived speed were higher for valid endogenous cues than neutral cues, although the significant main effect of the cue condition was observed only in the *easy* auditory discrimination task. In terms of task difficulty, the two-way repeated measures ANOVA test revealed a main effect of difficulty for temporal ventriloquism effect where the effect was higher in the hard difficulty than the easy difficulty as expected in the hypothesis.

As shown in Table 6.1, the difference between the hard experiment block and the easy experiment block was more prominent in the neutral cue case. This might be due to the fact that the increase in task difficulty forces the participant to attend more to the stimuli, thereby increasing the audiovisual processing. However, in the hard experiment block, the percent temporal ventriloquism effect difference between valid cues and neutral cues is less compared to the difference in the easy experiment block. This decrease might be due to the lack of resources left to process the cues in the hard experiment block since a big portion of the resources are used to complete the auditory discrimination task (Lavie, Hirst, Fockert, Viding, 2004). Considering this, Lavie and Dalton have proposed that distractors typically interfere more in a low perceptual load condition than a high perceptual load condition (Lavie, Dalton, 2014). However, the missed or unattended item could also work as another stimulus or a cue rather than a distractor, as in the case of our current study. In line with the perceptual load theories of attention, in a previous study, Calvillo and Jackson showed that inattention blindness resulting from high perceptual load occurs at an early stage of processing (Calvillo, Jackson, 2014). The inability to perceive a visible stimulus caused by the high attentional load required for another task is called *inattentional blindness* (Murphy, Groeger, Greene, 2016; Chabris, Simons, 2010). Moreover, the inability to perceive a stimulus under a high perceptual load is not unique to the visual modality. Macdonald and Lavie have shown that the level of perceptual load also affects the ability to notice an auditory stimulus while attending to an unrelated task, which they called *inattentional deafness* (MacDonald, Lavie, 2011).

As expected by our original hypothesis, response times were longer for neutral cues than valid cues even though statistical tests did not reveal any significant main effect of the cue in either of the blocks with different auditory task difficulties. Additionally, as the hypothesis predicted, the response times were longer in the hard blocks, the experiment with the harder task, than in the easy blocks. However, again statistical tests did not reveal a significant effect of difficulty. One reason for this might be that the tasks were already very demanding because of the number and configuration of the response keys. Even though participants were introduced to the response keys and trained on how to use them correctly, after the sessions some participants displayed that they had a hard time remembering the keys and reported that they took more time pressing the key because of this confusion. Hence, the response times may not be a measure as accurate as temporal ventriloquism in this sense.

## 5.1 Future Directions

The behavioral results of the current study revealed a significant cueing effect and task difficulty effect. An interesting follow-up study could be investigating at which stage of cortical processing this attentional modulation takes place. Electroencephalography (EEG) is a non-invasive method commonly used to measure the electrical activity on the scalp that represents the macroscopic activity of the brain's surface by capturing the fluctuations from the ionic current in the neurons (Niedermeyer, Silva, 2004). Thus, this experiment can be adapted into an EEG study in order to see which cortical areas are involved in the relationship of endogenous attention and audiovisual processes in the temporal domain. This neuroimaging experiment will be a natural extension of the current behavioral study. The behavioral experiment was designed by taking potential artifacts into account. For instance, the response keys were designed to minimize potential motor response artifacts that could be observed in EEG signals.

Numerous studies have investigated the neural mechanisms underlying multisensory processes. Mounting evidence suggests these processes take place over cortical areas at distinct stages (e.g., Kaya, Kafaligonul, 2021). In their EEG study, Kaya and Kafalıgönül investigated the cortical processes underlying the effects of temporal ventriloquism on perceived visual speed. They identified three main clusters of electrodes where auditory timing takes place: the medial-parietal, centro-parietal and right frontal electrodes (Kaya, Kafaligonul, 2019). Hence, in a future EEG study, it will be interesting to examine the effects of cueing on the activities over these identified scalp sites.

Based on the current behavioral results, it will be interesting to reveal cortical sites over which the main effects of cueing and difficulty are observed. Such observed effects on the neutral activities will provide important evidence for the attention effects as we have done in the behavioral experiments taking the attention theories such as load theory into account.

As we mentioned in the second chapter, Bayesian Framework has been used to model multisensory integrations. Hence, the behavioral results from this current study could be used to help make a modified Bayesian model.

It may be also worthwhile to understand the effects of different button layouts on behavioral performance since some of the participants had a hard time remembering the meanings of the buttons. A different button layout may increase the reliability of the response time measure and maybe even increase the temporal ventriloquism effect since the participants do not have to allocate more resources for the meaning of each button.

# REFERENCES

Alais D., Burr D. (2019). Cue Combination Within a Bayesian Framework. In: Lee A., Wallace M., Coffin A., Popper A., Fay R. (eds) Multisensory Processes. Springer Handbook of Auditory Research, vol 68. Springer, Cham. https://doi.org/10.1007/978-3-030-10461-0_2

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009, August). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*. https://doi.org/10.1016/j.conb.2009.06.008

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. Journal of the Optical Society of America. A, Optics, image science, and vision, 20(7), 1391–1397. https://doi.org/10.1364/josaa.20.001391

Bayes, T., "An essay towards solving a problem in the doctrine of chances", Philosophical Transactions of the Royal Society, Vol. 53, pp 370-418, 1783

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review*, *5*(3), 482–489. https://doi.org/10.3758/BF03208826

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception and Psychophysics*, *62*(2), 321–332. https://doi.org/10.3758/BF03205552

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225–2234. https://doi.org/10.1111/j.1460-9568.2004.03670.x

Blurton, S. P., Greenlee, M. W., & Gondan, M. (2015). Cross-modal cueing in audiovisual spatial attention. *Attention, Perception, and Psychophysics*, *77*(7), 2356–2376. https://doi.org/10.3758/s13414-015-0920-7

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev.neuro.26.041002.131052

Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press. https://doi.org/10.1037/10037-000

Bruns, P., & Getzmann, S. (2008). Audiovisual influences on the perception of visual apparent motion: Exploring the effect of a single sound. *Acta Psychologica*, *129*(2), 273–283. https://doi.org/10.1016/j.actpsy.2008.08.002

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*. Oxford University Press. https://doi.org/10.1093/cercor/11.12.1110

Calvillo, D. P., & Jackson, R. E. (2014). Animacy, perceptual load, and inattentional blindness. *Psychonomic Bulletin and Review*, *21*(3), 670–675. https://doi.org/10.3758/s13423-013-0543-8

Campagne, D. M. (2021). The problem with communication stress from face masks. *Journal of Affective Disorders Reports*, *3*, 100069. https://doi.org/10.1016/j.jadr.2020.100069

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*. https://doi.org/10.1016/j.visres.2011.04.012

Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition*, *102*(3), 321–340. https://doi.org/10.1016/j.cognition.2006.01.002

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. Attention, perception & psychophysics, 75(5), 790–811. https://doi.org/10.3758/s13414-013-0475-4

The invisible gorilla: and other ways our intuitions deceive us. (2010). *Choice Reviews Online*, *48*(03), 48-1757-48–1757. https://doi.org/10.5860/choice.48-1757

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

De Araujo, I. E., & Simon, S. A. (2009). The gustatory cortex and multisensory integration. *International Journal of Obesity*, *33*, S34–S43. https://doi.org/10.1038/ijo.2009.70

Denison, R. N., Carrasco, M., & Heeger, D. J. (2021). A dynamic normalization model of temporal attention. *Nature Human Behaviour*, *5*(12), 1674–1685. https://doi.org/10.1038/s41562-021-01129-1

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80–90. https://doi.org/10.1037/h0039515

Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience*, *9*(APR). https://doi.org/10.3389/fnint.2015.00032

Freeman, E., & Driver, J. (2008). Direction of Visual Apparent Motion Driven Solely by Timing of a Static Sound. *Current Biology*, *18*(16), 1262–1266. https://doi.org/10.1016/j.cub.2008.07.066

Grassi, M., & Casco, C. (2009). Audiovisual Bounce-Inducing Effect: Attention Alone Does Not Explain Why the Discs Are Bouncing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 235–243. https://doi.org/10.1037/a0013031

Hartcher-O'Brien, J., & Alais, D. (2011). Temporal ventriloquism in a purely temporal context. Journal of Experimental Psychology: Human Perception and Performance, 37(5), 1383.

Hirst, R. J., McGovern, D. P., Setti, A., Shams, L., & Newell, F. N. (2020, November 1). What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2020.09.006

IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0.
Armonk, NY: IBM Corp

Johnson, M. L., Palmer, J., Moore, C. M., & Boynton, G. M. (2020). Endogenous cueing effects for detection can be accounted for by a decision model of selective attention. *Psychonomic Bulletin and Review*, *27*(2), 315–321. https://doi.org/10.3758/s13423-019-01698-3

Jones, C. R. G., Happé, F., Baird, G., Simonoff, E., Marsden, A. J. S., Tregay, J., … Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, *47*(13), 2850–2858. https://doi.org/10.1016/j.neuropsychologia.2009.06.015

Kafaligonul, H., & Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals. *Journal of Vision*, *10*(12), 1–13. https://doi.org/10.1167/10.12.31

Kafaligonul, H., & Oluk, C. (2015). Audiovisual associations alter the perception of low-level visual motion. *Frontiers in Integrative Neuroscience*, *9*(MAR). https://doi.org/10.3389/fnint.2015.00026

Kaya, U., & Kafaligonul, H. (2019). Cortical processes underlying the effects of static sound timing on perceived visual speed. *NeuroImage*, *199*, 194–205. https://doi.org/10.1016/j.neuroimage.2019.05.062

Kaya, U., & Kafaligonul, H. (2021). Audiovisual interactions in speeded discrimination of a visual event. *Psychophysiology*, *58*(4). https://doi.org/10.1111/psyp.13777

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

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197. https://doi.org/10.3758/BF03213897

Lavie, N. (1995). Perceptual Load as a Necessary Condition for Selective Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2004.12.004

Lavie, N., & Dalton, P. (2014). Load Theory of Attention and Cognitive Control. *The Oxford Handbook of Attention*, (June), 56–75. https://doi.org/10.1093/oxfordhb/9780199675111.013.003

Loeffler, J., Cañal-Bruland, R., Schroeger, A., Walter Tolentino-Castro, J., & Raab, M. (2018). Interrelations between temporal and spatial cognition: The role of modality-specific processing. *Frontiers in Psychology*. Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2018.02609

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, and Psychophysics*, *73*(6), 1780–1789. https://doi.org/10.3758/s13414-011-0144-4

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

Mheidly, N., Fares, M. Y., Zalzale, H., & Fares, J. (2020). Effect of Face Masks on Interpersonal Communication During the COVID-19 Pandemic. *Frontiers in Public Health*, *8*. https://doi.org/10.3389/fpubh.2020.582191

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, *27*(15), 4120–4131. https://doi.org/10.1523/JNEUROSCI.4912-06.2007

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. In *Cognitive Brain Research* (Vol. 14, pp. 115–128). https://doi.org/10.1016/S0926-6410(02)00066-6

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163. https://doi.org/10.1016/S0926-6410(03)00089-2

Mukai, I., Bahadur, K., Kesavabhotla, K., & Ungerleider, L. G. (2011). Exogenous and endogenous attention during perceptual learning differentially affect post-training target thresholds. *Journal of Vision*, *11*(1), 25–25. https://doi.org/10.1167/11.1.25

Murray, M.M., Lewkowicz, D.J., Amedi, A., Wallace, M.T., 2016. “Multisensory processes: a balancing act across the lifespan”, *Trends in neurosciences*, 39, 567-579.

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin and Review*, *23*(5), 1316–1340. https://doi.org/10.3758/s13423-015-0982-5

Niedermeyer, E., & Silva, F. H. L. D. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. *Lippincott Williams and Wilkins*.

Ogulmus, C., Karacaoglu, M., & Kafaligonul, H. (2018). Temporal ventriloquism along the path of apparent motion: speed perception under different spatial grouping principles. *Experimental Brain Research*, *236*(3), 629–643. https://doi.org/10.1007/s00221-017-5159-1

Pew, R. W., & Mavor, A. S. (1998). Modeling Human and Organizational Behavior: Application to Military Simulations. *National Academy of Sciences*, 172–202. Retrieved from http://www.nap.edu/catalog/6173.html

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). *Attended and unattended processing modes: the role of set for spatial location*. *Modes of perceiving and processing information*. (pp. 137–158).

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(32), 3–25. 10.1080/00335558008248231.

Postma, D. B. W., Smith, J., Pepping, G. J., van Andel, S., & Zaal, F. T. J. M. (2017). When a fly ball is out of reach: Catchability judgments are not based on optical acceleration cancelation. *Frontiers in Psychology*, *8*(APR). https://doi.org/10.3389/fpsyg.2017.00535

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature neuroscience*, *16*(9), 1170–1178. https://doi.org/10.1038/nn.3495

Ramachandran, V. S., & Anstis, S. M. (1986). The perception of apparent motion. *Scientific American*, *254*(6), 102–109. https://doi.org/10.1038/scientificamerican0686-102

Remijn, G. B., Ito, H., & Nakajima, Y. (2004). Audiovisual integration: An investigation of the “streaming-bouncing” phenomenon. In *Journal of Physiological Anthropology and Applied Human Science* (Vol. 23, pp. 243–247). https://doi.org/10.2114/jpa.23.243

Roberts, K. L., Summerfield, A. Q., & Hall, D. A. (2009). Covert Auditory Spatial Orienting: An Evaluation of the Spatial Relevance Hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1178–1191. https://doi.org/10.1037/a0014249

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. In *Brain Topography* (Vol. 21, pp. 185–192). https://doi.org/10.1007/s10548-009-0090-9

Santangelo V., Macaluso, E. "19 spatial attention and audiovisual processing." The new handbook of multisensory processes (2012).

Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2003). Comparison of two weighted integration models for the cueing task: Linear and likelihood. *Journal of Vision*, *3*(3), 209–229. https://doi.org/10.1167/3.3.3

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency, of the ventriloquism effect. *NeuroReport*, *12*(1), 7–10. https://doi.org/10.1097/00001756-200101220-00009

Spagna, A., Wu, T., Kim, K., & Fan, J. (2020). Supramodal executive control of attention: Evidence from unimodal and crossmodal dual conflict effects. *Cortex*, *133*, 266–276. https://doi.org/10.1016/j.cortex.2020.09.018

Spence, C., & Driver, J. (1996). Audiovisual Links in Endogenous Covert Spatial Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 1005–1030. https://doi.org/10.1037/0096-1523.22.4.1005

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

Talsma, D. (2015, March 26). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*. Frontiers Media S.A. https://doi.org/10.3389/fnint.2015.00019

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. Neuroscience and biobehavioral reviews, 61, 208–224. https://doi.org/10.1016/j.neubiorev.2015.11.002

Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the "ventriloquism effect." *Perceptual and Motor Skills*, *36*(3 (II)), 1171–1184. https://doi.org/10.2466/pms.1973.36.3c.1171

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*(JUL). https://doi.org/10.3389/fpsyg.2014.00725

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, *3*(6), 449–459. https://doi.org/10.1016/S0022-5371(64)80015-3

Umiltà, C. (2000). Visuospatial attention. In F. Boller, J. Grafman, & G. Rizzolatti (Eds.), *Handbook of neuropsychology: Sect 1: Introduction, Sect 2: Attention* (pp. 393–425). Elsevier Science Publishers B.V..

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014, March 19). Multisensory integration: Flexible use of general operations. *Neuron*. Cell Press. https://doi.org/10.1016/j.neuron.2014.02.044

Van Opstal, J. (2016). *The auditory system and human sound-localization behavior*. Academic Press.

Vernet, M., Japee, S., Lokey, S., Ahmed, S., Zachariou, V., & Ungerleider, L. G. (2019). Endogenous visuospatial attention increases visual awareness independent of visual discrimination sensitivity. *Neuropsychologia*, *128*, 297–304. https://doi.org/10.1016/j.neuropsychologia.2017.08.015

Vroomen, J., & De Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 513–518. https://doi.org/10.1037/0096-1523.30.3.513

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Wickens, C. D. (1980). The Structure of Attentional Resources. In *Attention and Performance VIII* (pp. 239–257).

Yantis, S., & Johnston, J. C. (1990). On the Locus of Visual Selection: Evidence From Focused Attention Tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 135–149. https://doi.org/10.1037/0096-1523.16.1.135

Yantis, S., & Nakama, T. (1998). Visual interactions in the path of apparent motion. *Nature Neuroscience*, *1*(6), 508–512. https://doi.org/10.1038/2226

# **TEZ İZİN FORMU /** THESIS PERMISSION FORM

**ENSTİTÜ /** INSTITUTE

**Fen Bilimleri Enstitüsü** / Graduate School of Natural and Applied Sciences ☐

**Sosyal Bilimler Enstitüsü** / Graduate School of Social Sciences ☐

**Uygulamalı Matematik Enstitüsü** / Graduate School of Applied Mathematics ☐

**Enformatik Enstitüsü** / Graduate School of Informatics ☒

**Deniz Bilimleri Enstitüsü** / Graduate School of Marine Sciences ☐

**YAZARIN /** AUTHOR

**Soyadı** / Surname : KAVAKLIOĞLU
**Adı** / Name : EFSUN
**Bölümü** / Department : COGNITIVE SCIENCE (BİLİŞSEL BİLİMLER)

**TEZİN ADI /** TITLE OF THE THESIS (**İngilizce** / English) : SPEED PERCEPTION IN MULTISENSORY PROFILES: WHAT IS THE EFFECT OF ATTENTION?

**TEZİN TÜRÜ /** DEGREE: **Yüksek Lisans** / Master ☒ **Doktora** / PhD ☐

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☒
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two year**. * ☐
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. * ☐

*** *Enstitü Yönetim Kurulu Kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.*
*A copy of the Decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

**Yazarın imzası** / Signature ............................ **Tarih** / Date 01.03.2022