NEURAL MECHANISMS UNDERLYING SUB-SECOND CROSSMODAL TIME PERCEPTION

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

UTKU KAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

JANUARY 2018

NEURAL MECHANISMS UNDERLYING
SUB-SECOND CROSSMODAL TIME PERCEPTION

Submitted by UTKU KAYA in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Cognitive Science Department, Middle East Technical University** by,

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

Prof. Dr. Cem Bozşahin
Head of Department, **Cognitive Science**

Assoc. Prof. Dr. Annette Hohenberger
Supervisor, **Cognitive Science Dept., METU**

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

Examining Committee Members:

Prof. Dr. Metehan Çiçek
Physiology Dept., Ankara University

Assoc. Prof. Dr. Annette Hohenberger
Cognitive Science Dept., METU

Assist. Prof. Dr. Cengiz Acartürk
Cognitive Science Dept., METU

Assoc. Prof. Dr. Hüseyin Boyacı
Psychology Dept., Bilkent University

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

Date: 23-01-2018

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 : **Utku Kaya**

Signature : ____________________

ABSTRACT

NEURAL MECHANISMS UNDERLYING SUB-SECOND CROSSMODAL TIME PERCEPTION

Kaya, Utku
Ph.D, Department of Cognitive Sciences
Supervisor: Assoc. Prof. Dr. Annette Hohenberger
Co-Supervisor: Assist. Prof. Dr. Hacı Hulusi Kafalıgönül

January 2018, 125 pages

The integration of information from different senses is central to our perception of the world including the fundamental attributes of space and time. Audiovisual interactions have been well studied in this context. Similar to the fact that visual stimuli can drive the perception of where a sound comes from (spatial ventriloquism), auditory stimuli can drive the perceived timing of visual events (temporal ventriloquism). These interactions are in accordance with the auditory system's superior temporal and the visual system's superior spatial resolution. This experiment aims to understand the neural mechanisms underlying temporal ventriloquism. To achieve this objective, we collected behavioral and electroencephalography (EEG) data in tandem with four experiments. In Experiment-1, we have investigated how a single auditory event changes the perceived timing of a single visual event by varying the temporal offset between auditory and visual events. Experiment-2 focused on the mechanisms underlying the auditory influences on the visual time interval perception. In Experiment-3, we studied the effect of temporal ventriloquism on the perception of apparent motion by altering the timing of visual events leading to the motion percept. In the last experiment, we investigated the cortical mechanisms involved in the time interval adaptation aftereffects on the visual apparent motion. Overall, these experiments have significant contributions to our understanding of auditory influences on visual timing and crossmodal time perception in general. More generally, this experiment attempts to bridge the gap between our unified multisensory representation of the external world and low-frequency brain oscillations within the context of temporal ventriloquism.

Keywords: EEG, time and time interval perception, speed perception, audiovisual processes, sub-second range

ÖZ

ÇOKLU-MODALİTELİ SANİYE-ALTI ZAMAN ALGISININ OSİLASYON MEKANİZMALARI

Kaya, Utku
Doktora, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Doç. Dr. Annette Hohenberger
Ortak Tez Yöneticisi: Yard. Doç. Dr. Hacı Hulusi Kafalıgönül

Ocak 2018, 125 sayfa

Farklı duyulardan gelen bilginin entegrasyonu, uzam ve zamanın temel nitelikleri ile birlikte dünyayı algılamamızda merkezi bir konudur. Bu bağlamda özellikle görsel-işitsel etkileşimler iyi çalışılmıştır. Görsel uyaranların sesin geldiği yer algısında belirleyici bir rol oynaması gibi (mekânsal vantrilok-etkisi), işitsel uyaranlar da görsel olayların algılanma zamanlamasını etkileyebilmektedir (zamansal vantrilok-etkisi). Bu etkileşimler, işitme sisteminin zamansal olarak ve görsel sisteminin uzamsal olarak üstün çözünürlüğü ile uyumludur. Mevcut çalışma zamansal vantrilok-etkisinin altında yatan sinirsel mekanizmaları anlamayı amaçlamaktadır. Bu amaca ulaşmak için dört deneyde davranışsal ve elektroansefalografi (EEG) verileri birlikte toplanmıştır. Deney-1'de, bir işitsel olayın, bir görsel olayın algılanma zamanlamasını -işitsel ve görsel olaylar arasındaki zamansal farkın etkisiyle nasıl değiştirdiğini araştırdık. Deney-2, görsel zaman aralığı algılaması üzerindeki işitsel etkilerin altında yatan mekanizmalar üzerine odaklanmıştır. Deney-3'te zamansal vantrilok-etkisinin, hareket algılamasına yol açan görsel olayların zamanlamasını değiştirerek, zahiri hareket algısı üzerindeki etkisini araştırdık. Son deneydeyse, zahiri hareket algısı üzerindeki zaman aralığı adaptasyonu sonrası etkilerin kortikal mekanizmalarını araştırdık. Genel olarak, bu deneylerin, görsel zamanlama üzerindeki işitsel etkiler konusunda ve modaliteler-arası zaman algılaması konusundaki anlayışımıza önemli katkıları vardır. Daha genel olarak, bu çalışma, dış dünyanın birleşik çoklu-duyu temsiliyle zamansal vantrilok-etkisi bağlamındaki düşük frekanslı beyin salınımları arasındaki boşluğu kapatmaya çalışmaktadır.

Anahtar Sözcükler: EEG, zaman ve zaman aralığı algısı, hız algısı, görsel-işitsel süreçler, saniye-altı zaman aralığı

DEDICATION

To the ones who loved a beauty and fought for it.

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to my scientific-life-long supervisor Dr. Annette Hohenberger for her encouraging support. I strongly believe that I would not be doing science without her support.

I would like to express my gratitude to my supervisor Dr. Hulusi Kafaligonul for his persistence to ask for a higher quality of the work, and for his support in all phases of this thesis from designing it to writing the manuscript. His kind friendship even in the days that things were not so bright for me was also so important to finish my thesis.

Besides my supervisors, I would like to thank my thesis committee members, Dr. Metehan Çiçek, Dr. Murat Perit Çakır and the jury members Dr. Cengiz Acartürk and Dr. Hüseyin Boyacı for their valuable comments and critics.

As a team, my dearest lab friends Sibel Akyüz, Fazilet Zeynep Yıldırım, Cansu Öğülmüş, Ayşenur Karaduman. Undergraduates Mert Özkan, Can Oluk, Deniz Duman, Toygan Kılıç, Yasemin Abra helped me in various steps of my thesis. Thank you all, you were awesome.

I would also like to thank all the colleagues from our kitchen table discussions in UMRAM: Ozan Emir, Mustafa Can Delikanlı, Zahide Pamir Karatok, Buse Merve Ürgen, Alp.

Beyond all, I should express my deep gratitude to my Family, Nilgün and the members of METU Players, those who gave me a peaceful home to sleep.

This work was supported by The Scientific and Technological Council of Turkey (TÜBİTAK 113K547).

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF ABBREVIATIONS

A1	Primary Auditory Cortex
AC	Auditory Cortex
AES	Anterior Ectosylvian Sulcus
ANOVA	Analyses of Variance
EEG	Electro-Encephalography
ERP	Event-Related Potential
fMRI	Functional Magnetic Resonance Imaging
ISI	Inter-Stimulus Interval
PFC	Prefrontal Cortex
PP	Posterior Parietal Cortex
ROI	Region of Interest
SOA	Stimulus Onset Asynchrony
SC	Superior Colliculus
STS	Superior Temporal Sulcus
V1	Primary Visual Cortex
VC	Visual Cortex

CHAPTER 1

1 INTRODUCTION

In our daily lives, we constantly interact with the external environment through perception and action cycle. Accuracy and speed are always very important during these interactions. An event in the external environment creates various forms of energy. Energy is transferred in the environment by different mediums. Hence, we are equipped to detect energy changes in different channels (i.e., different modalities of information). Such energy changes (i.e., sensory information) are transduced into neural signals by different sense organs (Fig. 1). Sensory information which is over a certain threshold is transduced into electrical signals. In our brain, those signals are integrated in such a way that unique perceptions of the events emerge. Therefore, various channels of information transduced from different modalities are integrated and combined. Combining streams of sensory information provided by external environment helps us to construct coherent multisensory percepts, and hence we interact and take necessary motor actions.

The perception of an event also implies that perception of some crucial properties of the event, such as timing and location. Information about those properties is transduced by the sensory organs and channels of information are processed in the various locations of the brain over their travel in the brain. Interestingly, however, information about some properties is mainly carried via more than one sensory system. For example, perceiving time and location of an event is extremely important, and information about these properties is carried on different sensory channels. For instance, when an event occurs, we locate the event only by hearing its sound or also locate it only by seeing it. The same concept applies to the perception of the time of the event. In other words, both auditory and visual systems carry spatial and temporal information. In fact, combining different modalities is obviously an advantage for organisms, because different modalities have different advantages in detecting specific properties. For example, it is very well known that audition has higher accuracy in time while vision is better at spatial resolution (Alais & Burr, 2004a, 2004b). However, due to the variations in physics of the energy transfer in nature and to the differences in the design of the biological mechanisms of the sensory organs, sensory channels may provide conflicting information. Interestingly, organisms have such compensating multisensory integration mechanisms that resolve such conflicts in the property information between modalities, to some extent.

Carrying a different kind of information about a property in distinct sensory channels generally does not mean that these information channels are conflicting, yet they are complementing each other and create a new unique perception of the property. Many multisensory integration paradigms provide illustrations of this view. For instance, in McGurk effect auditory and visual channels carry different information about the syllable; the perception of the syllable is qualitatively different than each mode of information (McGurk & Macdonald, 1976). Also, in both spatial and temporal ventriloquism effects, although the auditory and visual channels carry different information about location and time; perceptions of these properties are unique, which are a combination of both channels, although one channel dominates the other at a specific property(Chen & Vroomen, 2013).

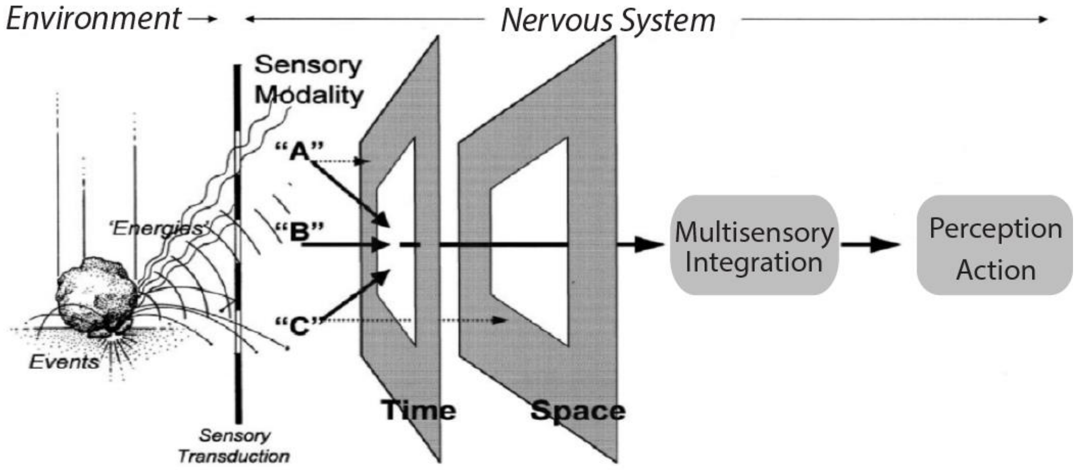


Fig. 1. Multisensory processing of the external world events. External events, such as a falling rock, create different forms of physical energies. Different sensory receptors transduce different forms of energies into neural signals. However, multisensory stimuli must occur within specific temporal and spatial windows, to get interact in the brain. Since different modalities of sensory data are processed in various aspects such as time and space, events that occur outside these windows (dashed arrows) are processed as separate, hence perceived as modality-specific events (adapted from Meredith, 2002).

Considering the notable audiovisual integration mechanisms, it is not surprising that such interactions have very fundamental neural structures in our brain. But, the yet unclear part of this interaction is where and when, and how the information is integrated into the brain. In the recent literature, contrary to older studies, sensory processing and perception are not viewed in isolation from different sensory modalities, operating largely independently. However, new studies of multisensory integration in various perceptual tasks and settings suggest that crossmodal interactions also occur within the early sensory processing areas which are thought to be "sensory-specific" (Ghazanfar & Schroeder, 2006; Kayser & Logothetis, 2007; Kayser *et al.*, 2008). These findings suggesting that multisensory processing starts at relatively early information processing locations. But, in terms of the sensory interactions in the temporal dimension, signals come at different times and they travel through the brain and it is unlikely that they can interact with each other at the same spatiotemporal profile. Although the neural structures where different modalities interact at are the same, travel of the signals has a temporal dimension. Therefore, it also affects the temporal dimension of the interaction mechanisms. In other words, signals with different SOAs may interact at different locations at different times. To sum up, it is very important to have a good understanding of the neural mechanisms of the temporal integration of multisensory information.

1.1 Specific Aims and Research Questions

This dissertation, in general, aims to contribute to our understanding of the neural mechanisms underlying crossmodal interactions. Specifically, audiovisual interactions in time and their perceptual consequences were investigated. From scope of the temporal ventriloquism and multisensory integration in general, my focus is on the time dimension of the visual event perception which includes perception of a single event timing (earlier or later), perception of time interval between two consequent events (shorter or longer) and speed perceptions of two events at different locations (faster or slower). As auditory modality has higher resolution in the time dimension, the temporal structure of the concurrent and pre-stimulus presentation of auditory stimuli affect these perceptual qualities.

Therefore, the main aim of this dissertation is to investigate the neural mechanisms underlying temporal ventriloquism, and the influences of the auditory temporal structure on time interval perception. Additionally, influences of both concurrent and pre-test auditory temporal structure on motion perception were investigated. We tested the following two hypotheses to achieve our specific aims. First, we wanted to determine the stages and levels of audiovisual interactions in time, with a specific question whether these crossmodal interactions occur at early levels of perceptual processing or not. To determine the neural correlates of auditory modulation on time interval and speed perception, we examined low-frequency EEG activity recorded over different parts of the sensory cortices. The idea is that crossmodal interactions are not restricted to the high associative areas, but the interactions can also be observed at the early levels of perceptual processing.

This hypothesis originates from the latest studies that show strong evidence for early crossmodal interaction (Kayser & Logothetis, 2007). Due to a possible direct link between the auditory and visual cortices, activation in the early auditory cortex might cause an additional activity and modulation on to activity at early visual cortices. Therefore, unified excitement in the early visual cortex initiates a unified perception of different modalities at a very early processing step. This hypothesis was tested in four experiments: In the first experiment, simple single clicks single flash experiment, where varying onsets of the different modalities were investigated both in behavioral and especially in brain oscillations level. Like the first experiment, time interval and apparent motion perception under the influence of auditory stimulus in the second and third experiments, respectively, were also tested.

In the last experiment, influences of the auditory and visual time interval adaptation on the apparent motion were investigated.

The second specific aim of this dissertation is to elucidate the oscillatory patterns underlying shifts in the timing of visual perception. In addition to the neuroanatomical basis of multi-sensory integration, the functional significance of oscillatory activity for crossmodal interactions and multisensory integration are still subject to debate. Accordingly, the second main hypothesis is that auditory timing modulates the phase, duration and/or magnitude of induced neural responses to the visual event, time interval, and apparent motion. This hypothesis follows from the assumption that the shifts at perceived visual timing and speed result from the changes in the specific character of the EEG signals such as phase, amplitude, frequencies, and duration of neural responses to the visual time intervals and apparent motion. These neural changes responding to visual stimuli were also modulated according to the timing of auditory stimuli at the brain areas where the sensory processing pathways cross. We will test this hypothesis by comparing the response duration for different auditory timing conditions. Alternatively, the perceived visual interval and changes in perceived visual speed might be observed in changes in the amplitude of various frequencies of EEG oscillations especially over occipital areas, both at early feedforward and late feedback processing steps.

1.2 Organization of the Dissertation

This dissertation is organized as a distinct chapter. In Chapter 2, literature is reviewed. First, audiovisual integration is reviewed as a phenomenon with the help of a few illusions at the behavioral level. This chapter discusses the possible neural mechanisms and reviews related neuroimaging studies that were designed for testing the main hypotheses in this dissertation. Following four chapters present these four empirical studies. First, auditory effects on the time perception of a visual event are investigated (Experiment 1) in Chapter 3. In this experiment, participants are exposed to only one flash and their behavioral and neural responses are recorded and analyzed. Chapter 4 discusses the time interval judgment experiment (Experiment 2) where subjects are exposed to visual time intervals that are manipulated by the concurrent auditory stimulus. In a highly similar experiment, the visual apparent motion is modulated by auditory time intervals (Experiment 3) which are reported in Chapter 5. Although apparent motion is modulated by shorter and longer time intervals of concurrent auditory stimulus, in the Chapter 6 influences of pre-stimulus time interval adaptation on visual apparent motion are investigated (Experiment 4). Time intervals that are used for adaptation in this experiment are defined both in visual and auditory modalities. Therefore, this modality effects on speed perception are also investigated. In conclusion, Chapter 7 discusses the finding of this thesis in the scope of the neural mechanisms of audiovisual temporal processing and concludes the general findings of this dissertation.

CHAPTER 2

2 LITERATURE REVIEW

2.1. Origins of multisensory research

As organisms are adapted to the environment which has sources of many forms of energy they have mechanisms for gathering these sources information from the environment. The transduction of these different forms of energy as inputs to the cognitive and perceptual system has a clear adaptive advantage for organisms. Therefore, many organisms are adapted to have various sense organs to transduce the energy changes in the environment. For instance, fast animals should have such neural mechanisms providing a robust and fast perception of the events in the environment, so they can act accordingly. Such a mechanism has a clear survival importance. Perceiving various aspects of environment requires deducing the information from different forms of energy with various sensory organs. Therefore, the information coming from different modalities must be integrated in a way that they are perceived as they are coming from one source at a specific time. Beyond all other senses, auditory and visual systems have a major role to perceive the events at a distance with a considerable speed. Due to the differences in the physical energy transfer mechanisms between sound and light waves, depending on the distance, a time difference is introduced till they reach the sensory organs. However, it is sometimes vital to integrate the two modalities as they help to perceive an event at a distance. As mentioned above, speed and accuracy are vital in perception in perception. Therefore, auditory signals may help in speed and accuracy in the perception of the visual event, while the other way around is also correct; and possible mechanisms which are dedicated to it are very intriguing.

In general, we know that audiovisual interactions illustrate the more general phenomenon that different modalities interact and that they change the perception of each other (Calvert *et al.*, 2004). It is a matter of fact that multisensory integration is a fundamental aspect of cognition and integration is very important for a coherent perception (Mazaheri *et al.*, 2013). Moreover, many studies showed that multisensory input can facilitate behavior by speeding reactions up (Hershenson, 1962; Gielen *et al.*, 1983). Multisensory input can also improve detection of faint stimuli (Frens & Van Opstal, 1995; Driver & Spence, 1998; McDonald *et al.*, 2000; Vroomen & de Gelder, 2000; Feng *et al.*, 2014). Different sensory mechanisms do not affect perception only in quantity, but changes the perception also qualitatively as in the illusions such as ventriloquism, the McGurk effect, the stream-bounce (Howard & Templeton, 1966; McGurk & Macdonald, 1976; Sekuler *et al.*, 1997; Jousmäki *et al.*, 1998; Shams *et al.*, 2000; Guest *et al.*, 2002). Such an important function should be actualized with some substantial neural structures in the brain.

2.1 Audiovisual interactions

Audiovisual interactions give us a good framework to understand crossmodal interactions and multisensory integration. Timing and locations of the events are two important properties of the events that are perceived via information coming from both auditory and visual sensory organs independently. Both sensory systems are capable of detecting each property to some extent with no entrainment of the other, however, when information comes from both channels they affect each other and create a unique perception of timing and location (Recanzone, 2009). Sound-induced flash illusion, spatial ventriloquism, and temporal ventriloquism are good examples of audiovisual interactions, which are specifically related to the perception of timing and location.

2.1.1 Sound-induced flash illusion

A strong approach to study crossmodal interaction is to use the same physical sensory events that can lead to different percepts with the effect of another sensory stimulus. The sound-induced flash illusion is a well-known example in which participants are exposed to one flash accompanied by multiple beep sounds. As a result, based on the number of beeps used, observers typically perceive two illusory flashes or one non-illusory flash (see Fig. 2.a) (Shams *et al.*, 2000, 2002). When trials, where participants reported two illusory and one non-illusory percept, were compared, neural correlates of the perceptual difference can be observed in the early visual cortex. Event-related potentials and oscillatory mechanisms of this illusion is elucidated and they show that second illusory flash is based on a very rapid interaction between the visual and auditory cortices which are triggered by the second beep (Shams *et al.*, 2005; Mishra *et al.*, 2007, 2008) (see Fig. 2B).

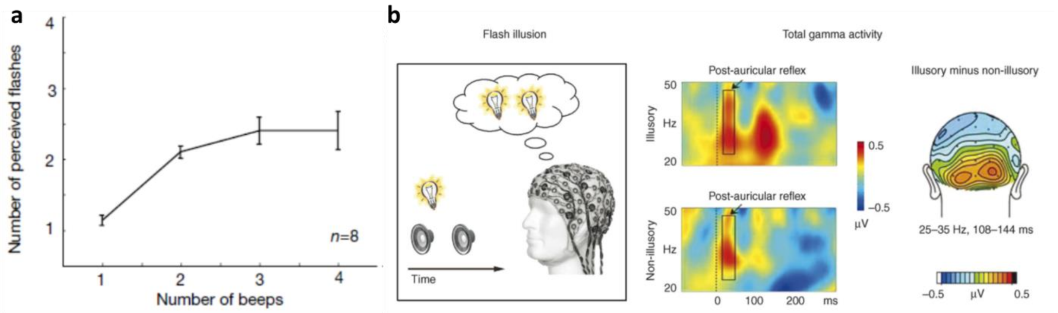


Fig. 2. Illustration of the sound-induced flash illusion. (A) Illusory flashing. The average number of perceived flashes is plotted against the number of beeps, for trials in which the visual stimulus consisted of one single flash. Observers report seeing two or more flashes when the single flash is accompanied by two or more beeps (adapted from Shams et al., 2002) (B) Gamma-band activity

during a flash illusion experiment. A single flash that is presented interposed between rapidly occurring auditory inputs is frequently perceived as multiple flashes. Gamma-band-activity was enhanced over occipital scalp in a time window between 100 and 150 ms during illusory trials in which multiple flashes were perceived compared to non-illusory trials in which only a single flash was perceived (adapted from Mishra et al., 2007).

2.1.2 Spatial and temporal ventriloquism

If auditory and visual stimuli occur close enough in time and space, but not necessarily at the same time and location, they typically perceived as originating from one unified event. This illusion is named as ventriloquism, coming from Latin for "*to speak from the stomach*" (i.e. *venter*: belly and *loqui*: speak). Ventriloquism is largely known as the illusion about the location where the sound originates. In case of being close in space, visual stimulus captures the location of the auditory one, in accordance with the superiority of visual system in spatial resolution (*spatial ventriloquism*) (Howard & Templeton, 1966; Bertelson & Aschersleben, 1998; Alais & Burr, 2004a) (see Fig. 3A). If two sensory modalities come close in time but within an onset asynchrony, auditory stimulus captures the perceived timing of the visual event, which makes an adaptive sense with the superiority of auditory system in time domain (*temporal ventriloquism*) (Fendrich & Corballis, 2001; Morein-Zamir *et al.*, 2003; Recanzone, 2003; Vidal *et al.*, 2017) (see Fig. 3B). Temporal ventriloquism can also be observed in such a case that perception of the visual time interval is also modulated by the time interval marked by brief sounds such as clicks. (Scheier *et al.*, 1999) (see Fig. 3C).

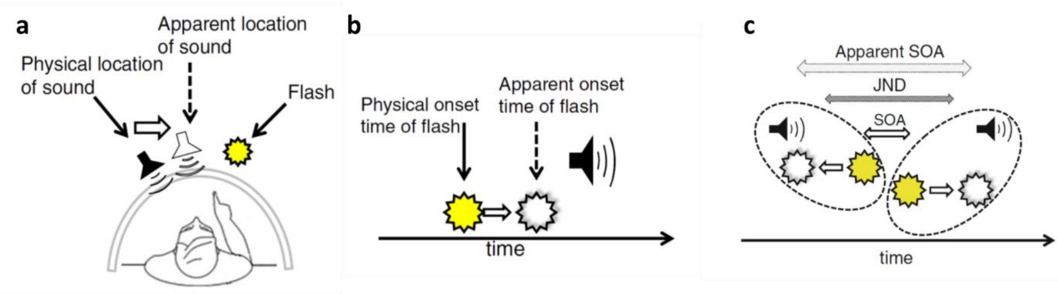


Fig. 3. Illustration of the spatial and temporal ventriloquisms. (A) Spatial ventriloquism: The apparent location of an auditory target sound is shifted in the direction of a spatially displaced visual stimulus (a flash). (B) Temporal ventriloquism: The apparent onset of a flash is shifted toward a sound that is presented at a slightly different timing than the flash. (C) Observers judge which of two flashes (upper or lower) appeared first. The judgment of visual temporal order is difficult because the flashes are presented at a stimulus onset asynchrony (SOA) below the just noticeable difference (JND). Two click sounds, one just before the first flash and the other after the second flash, make this task easier because the clicks shift the apparent onsets of the flashes, thereby increasing the apparent SOA above the JND. Sensitivity for visual temporal order thus improves when flashes are "sandwiched" by clicks (AVVA). If the clicks were presented in between the two flashes (VAAV), sensitivity would become worse, because clicks then decrease the apparent SOA (Scheier et al., 1999) (images are adapted from Chen & Vroomen, 2013).

Temporal ventriloquist effect is also observed in motion perception. Motion depends on spatiotemporal correlation which depends on spatial and temporal information provided by each frame. As we know from temporal ventriloquism, temporal component of the visual event can be determined by different modalities. Therefore, when the temporal component is modulated by audition, the temporal structure of audition also affects speed perception of the visual motion. Many studies have shown that direction of the motion perception in visual modality can be modulated by the auditory modality that does not have spatial information but carries high-resolution temporal information (Freeman & Driver, 2008; Kafaligonul & Stoner, 2010; Shi *et al.*, 2010). Several studies have also reported that the timing of brief sounds can alter how visual AM stimuli are perceived, presumably by altering the perceived timing of the visual stimuli that yield the motion percept (Staal & Donderi, 1983; Getzmann, 2007; Freeman & Driver, 2008; Kafaligonul & Stoner, 2010). For example, if one click precedes the first visual frame by 20 ms and the second click follows the second visual frame by 20 ms, the perceived interval is lengthened and perceived speed decreases (Freeman & Driver, 2008). Besides being a striking effect, this multisensory interaction is a powerful tool to investigate the multisensory dynamics of time perception.

Although there are many behavioral investigations in the audiovisual interactions, neural mechanisms underlying temporal ventriloquism are not completely understood and neural correlates are unknown. Several accounts can be hypothesized on the nature of audiovisual interactions. One possible candidate is phase resetting whereby phase reset of ongoing neural oscillations modulates detection of a sensory target (Naue *et al.*, 2011; Thorne *et al.*, 2011; Thorne & Debener, 2014). It is suggested that stimuli in one modality get multisensory cortical regions more prone to be excited. When another modality stimulates the same area, multisensory areas get excited in a way that can be observed in a phase alignment.

2.2 The hierarchical model

A fundamental question in multisensory research is how channels of information provided from different modalities are combined. There are several views in the literature about the processing stages of multisensory integration. Two opponent views can be put forward. One of them proposes that unisensory information is first processed at early cortical areas and travels into the higher association areas where various sensory signals interact and they are integrated. These brain regions are located both at the cortical and subcortical regions. Superior colliculus, superior temporal sulcus, the intraparietal sulcus and regions in the frontal lobe (see Fig. 4) are mostly mentioned areas that are known as responsible for association and there are many functional and anatomical studies that support crossmodal interactions in these areas (Jones & Powell, 1970; Felleman & Van Essen, 1991; Stein & Meredith, 1993; Kayser & Logothetis, 2007). There are various both functional and anatomical studies supporting this view (Benevento *et al.*, 1977; Bruce *et al.*, 1981; Hikosaka *et al.*, 1988; Graziano *et al.*, 1994, 1999; Calvert *et al.*, 2000; Fuster *et al.*, 2000;

Beauchamp *et al.*, 2004; Barraclough* *et al.*, 2005; Saito *et al.*, 2005; Sugihara *et al.*, 2006; Avillac *et al.*, 2007; van Atteveldt *et al.*, 2014).

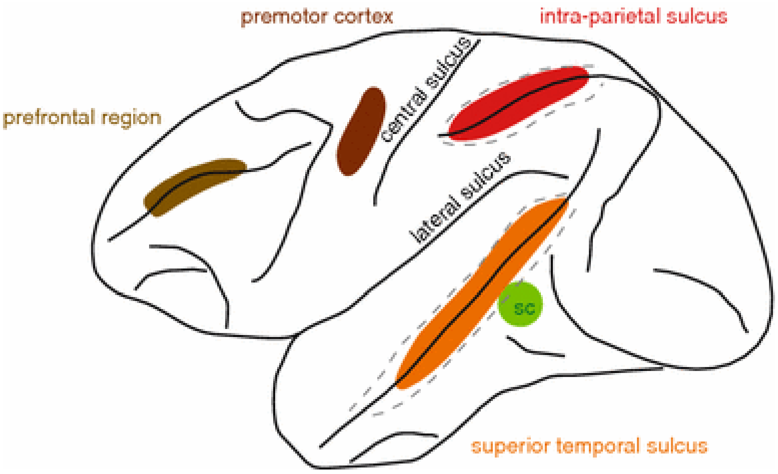


Fig. 4. Association areas implied in sensory integration. The (subcortical) superior colliculus (SC) is shown in light colors and the dashed gray lines indicate regions where sulci were "opened". See text for a list of references reporting sensory integration in these areas. Adapted from (Kayser & Logothetis, 2007).

2.3 Models primarily depend on low-level interactions

As mentioned in previous sections, the classical view claims that multisensory processing occurs in a hierarchical manner throughout convergence in neural pathways. According to this view, the signal coming from different senses merges at higher association areas, since subcortical and early cortices are specialized for unimodal processing (Meredith, 2002). The core idea of this approach is that signals from different senses are represented by the firing rates of the neurons. Group of neurons from one modality may have an excitatory or inhibitory effect on the others at conventional multisensory areas. Hence, the quality of the final perception is determined by the quantities of each interacting modality (Senkowski *et al.*, 2008). Although there have been older studies suggesting this perspective for a long time (Murata *et al.*, 1965; Bental *et al.*, 1968; Spinelli *et al.*, 1968; Morrell, 1972; Fishman & Michael, 1973; Vaudano *et al.*, 1991), early sensory integration models, on the contrary side, claiming that pure convergence approach does not explain all the aspects of crossmodal interaction, had more attention recently (Shimojo & Shams, 2001; Schroeder *et al.*, 2004; Foxe & Schroeder, 2005; Ghazanfar & Schroeder, 2006; Kayser & Logothetis, 2007; McDonald *et al.*, 2013).

Reasons suggesting this approach have been grouped into three categories. First, there is very strong experimental evidence showing that strong crossmodal interactions occur already in the primary cortices, which hierarchical convergence view cannot account for (Ghazanfar & Schroeder, 2006; Kayser & Logothetis, 2007). Second, there is strong experimental evidence showing that multisensory integration can account for the fast changes in the multisensory data. Third, convergence is a feedforward mechanism where high-level representations do not have elementary properties since convergence model does not explain how percept still has low-level information about the objects (Senkowski *et al.*, 2008).

2.4 Anatomical pathways and multisensory areas

Recent studies show that the adult male human brain contains on average 86 billion neurons and 85 billion nonneuronal cells (Herculano-Houzel, 2009). Almost every part of them is connected to others with synaptic connections. Signals are transferred through the neurons if once the neuron is fired. The probability of being fired is related to be the frequencies of presynaptic neurons and the excitation threshold of the cell body. Therefore, a signal must pass from various borders to be transferred from one location to another. Thus, the structure of the brain predefines the channels and the borders or thresholds that a signal can be transmitted. Naturally, anatomical structures give a perspective about how information is transferred. In fact, for such a phenomenon of multisensory integration, the pathways are more important for that unisensory information comes from different sources and enters the brain in different locations. As suggested by early sensory integration models, integration should occur fast and robustly via direct pathways connecting early sensory processing areas (Driver & Noesselt, 2008).

There are many different anatomical regions whose multisensory integration roles emphasized in the literature. Superior colliculus (SC) is one of the very well-known of these areas, which is located at the midbrain. SC contains neurons carrying information from auditory, visual and somatosensory modalities (Stein & Arigbede, 1972) and widely known for its role in the control of rapid eye and head movements which are highly related with the information coming from the visual and auditory sensory modalities. Interestingly, this area, in terms of anatomical location and function, corresponds to the optic tectum in avians and reptiles, whose proportion to the cortex is much bigger.

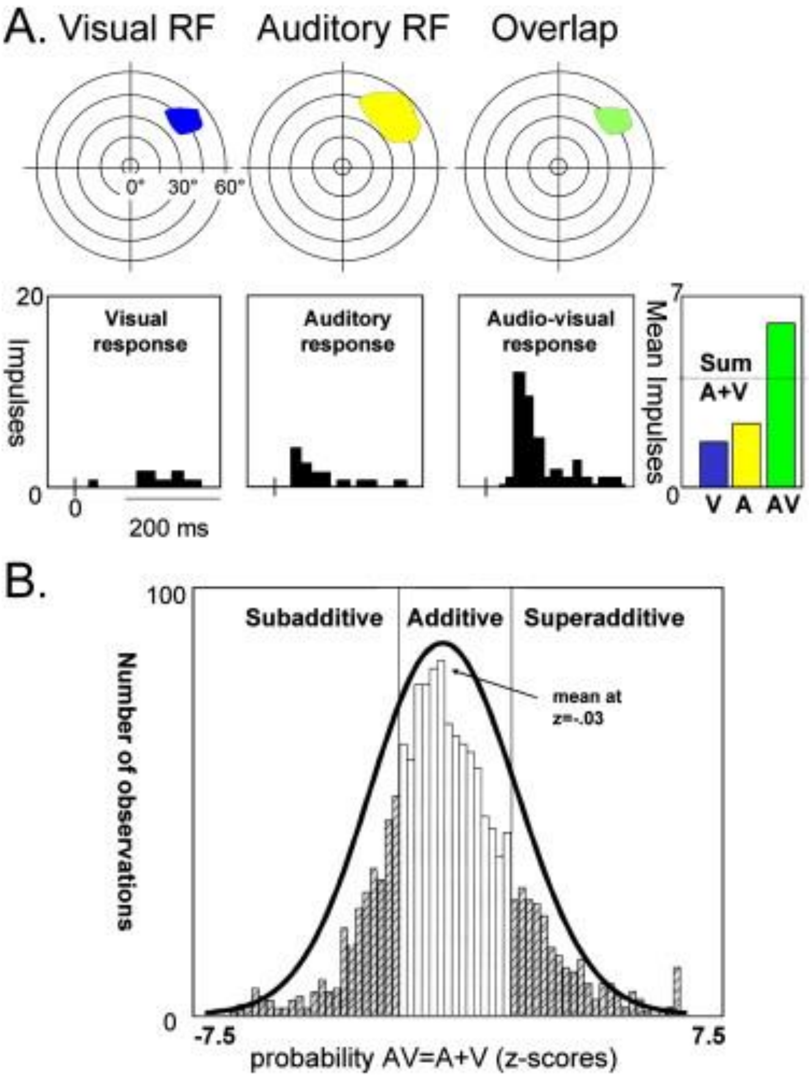
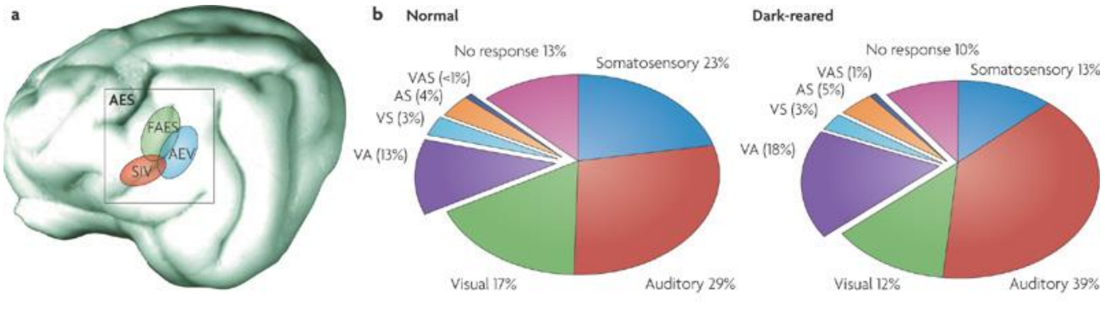


Fig. 5. Response properties of multisensory neurons. (A) Multisensory neurons in deep SC of the cat brain show super-additive pattern. As observed, when the auditory and visual stimuli spatiotemporally overlapped evoked responses exceeds the summation of unisensory evoked responses. (B) Distribution of z scores for a population of sampled neurons within deep layers of the cat superior colliculus, where z scores relate to firing rates for combined audiovisual stimulation, as compared with summed unisensory auditory and unisensory visual responses. Although the responses predominantly show an additive pattern some of the neurons may also show super and sub-additive patterns (adapted from Driver & Noesselt, 2008).

Various levels of neural measurements in multisensory research provide a wide range of information. Single multisensory neurons that respond to multiple sensory data are defined as multisensory neurons. Single neuron measurements based on the cells from *superior colliculus* is one of the prominent research areas. These single neuron studies derive a number of principles about sensory integration. The first principle is the *spatial coincidence*. According to this principle, multisensory neurons respond only when information about the location of the stimuli in different sensory channels overlaps (Stein, 1998). The second principle of *temporal coincidence* poses that only the stimuli that occur in a temporal proximity enhance the neural activity (Stein & Wallace, 1996). Together with the first principle, neurons that are selective only for the same event are identified as multisensory integration. Having a specific time and location range indicates that there is no 'any' kind of neural architecture, but rather, some 'specific' kind of connection should also be observed. The third principle is *inverse effectiveness*. Generally, a unisensory stimulus that elicits a strong response does not interact with information in other sensory modalities, but a stimulus that elicits weak response may cause a strong response due to interaction with another sensory stimulus when presented in spatial and temporal proximity (Perrault *et al.*, 2003). This last principle in single neuron level has a nice link between a neural activity and behavioral benefits of multisensory integration. When unisensory information is strong enough, it elicits a strong response, but when a sensory stimulus is weak, that stimulus may elicit an even stronger response with the enhancement of the other weak sensory stimulus (see Fig. 5).

There are some other parts which are highlighted, too, for multisensory integration. For example, studies that are done with cats, the anterior ectosylvian sulcus, which contains neurons sensitive to unisensory visual, auditory, and somatosensory stimulus, is also sensitive to multisensory stimuli (see Fig.6). The functional roles of these areas have been changed when those sensory processing mechanisms have been deprived. Another example is that the occipital cortex gets activated when somatosensory and auditory activities such as braille reading and tactile discrimination and also auditory localization and discrimination take place as shown in Fig. 7a (Merabet & Pascual-Leone, 2010). Similarly, the auditory and language-related areas are also being used for visual and tactile processes (see Fig. 7b). This is also strong evidence of the phenomenon that visual and auditory cortices of the brain regions are very closely related to multisensory processing, besides the well-known multisensory areas. Multisensory integration can be observed both at the single neuronal level and at groups of the neuronal level at various cortical regions (Meredith *et al.*, 2012). When a single neuron was observed, for example at SC, it was observed that there is an immediate convergence at real time (Miller *et al.*, 2017).



Nature Reviews | Neuroscience

Fig. 6. Dark rearing alters the distribution of sensory responsive neurons in the anterior ectosylvian sulcus (AES). (A) The lateral surface of the adult cat cortex, with the location of the AES and the relative positions of its three major subdivisions: SIV (fourth somatosensory area), FAES (auditory field of the AES) and AEV (anterior ectosylvian visual area). (B) The distribution of sensory unresponsive, unisensory and multisensory mature AES neurons in normally reared and dark-reared animals. Values are rounded to the nearest percent. AS, auditory-somatosensory; VA, visual-auditory; VAS, visual-auditory-somatosensory; VS, visual-somatosensory (Adapted from Merabet & Pascual-Leone, 2010).

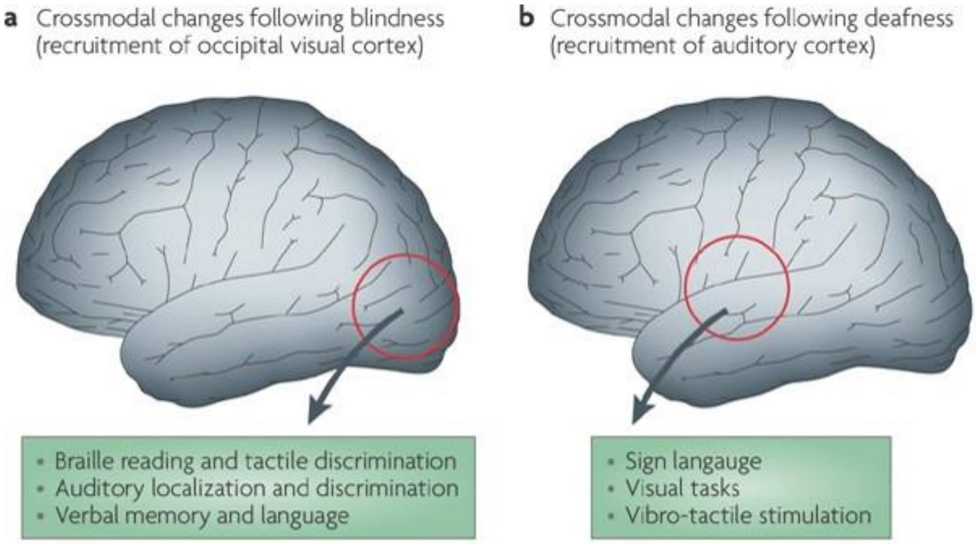


Fig. 7. Crossmodal recruitment of occipital visual cortex in the blind and auditory cortex in the deaf has been reported. (A) Occipital recruitment for tactile processing such as Braille reading, sound localization, and verbal memory. (B) Recruitment of auditory and language-related areas for viewing sign language, peripheral visual processing and vibrotactile stimulation (Adapted from Merabet & Pascual-Leone, 2010).

To sum up, multisensory integration is processed in many locations in both brain stem and cortex. There is profound evidence supporting that SC is the first location that the multisensory interactions take place (Fig. 8A.I). Both integrated and unimodal signals are distributed to primary cortices, and multisensory interactions are also occurred in between primary sensory cortices (Fig. 8A.II). And the signals are transmitted and integrated also in conventionally known as multisensory regions such as the one which is close to primary auditory cortex (Fig. 8B). Besides the feed-forward interactions, the signals integrated into some multisensory regions have feedback connections to the regions that are unisensory and the unisensory signal is also integrated with these feedback signals (Fig. 8C). For a specific task where multisensory integrations occur, some or all of these integrations mechanisms may play a role in the integration at the same time.

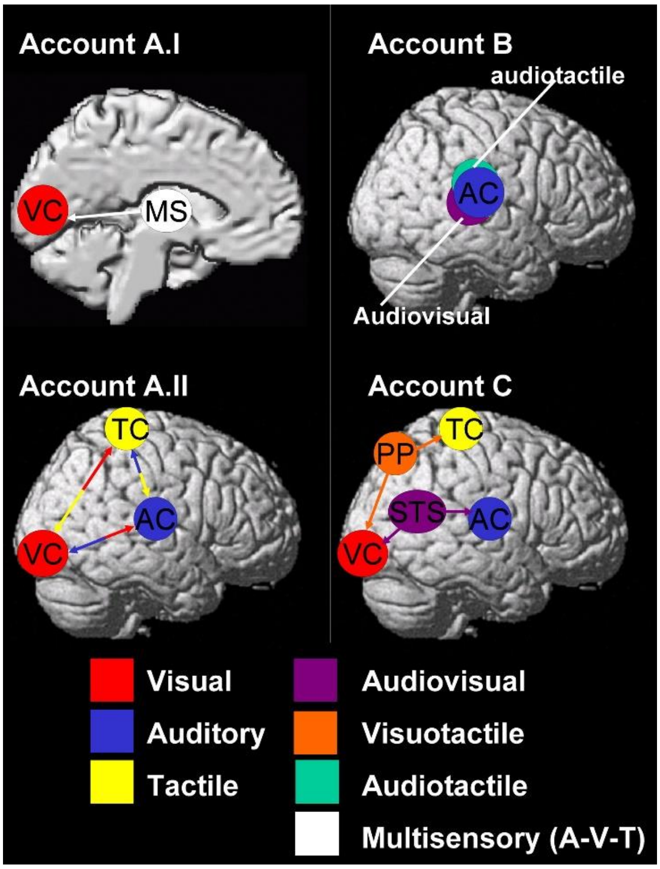


Fig. 8. Possible neural pathways mediating multisensory interplay. (A) Direct feedforward influences between visual and auditory processing, which might either arise subcortically at thalamic levels, as sketched in (I), if multisensory (MS) thalamus influences visual cortex (VC); and/or via sparse cortical-cortical connections directly between auditory cortex (AC, blue), visual cortex (VC, red), and somatosensory or tactile cortex (TC, yellow), as in (II). (B) Some multisensory regions may exist near classic unisensory regions, as for some audiovisual areas (violet) and some audio-tactile (green) areas near conventional auditory cortex (blue). (C) Feedback connections may exist between higher-level multisensory regions and back to lower-level areas that are (predominantly) sensory-specific apart from these feedback influences. For instance, visual and tactile modalities may interact via particular regions of posterior parietal cortex (PP, orange) that receive afferent input from both modalities and send feedback projections to each; and analogously, auditory and visual modalities may interact in posterior STS (violet) and send feedback projections to sensory-specific auditory and visual cortex (adapted from Driver & Noesselt, 2008).

2.5 Role of EEG Methodology in Multisensory Research

Electroencephalography (EEG) provides great temporal resolution in terms of the neural activity captured on the scalp. Due to the principle of the temporal coincidence, multisensory stimulus should be close in time which is around 200-ms time window. Neural dynamics within such short timescales can successfully be accounted by EEG. The amplitude and phase of these oscillations have been shown to play an important role in understanding the dynamics of crossmodal interactions. For instance, the influence of an auditory click on the sensitivity to a briefly presented visual flash is reflected in amplitude and phase changes in EEG frequencies (10-30 Hz) over visual cortex (Naue *et al.*, 2011; Fiebelkorn *et al.*, 2013; Mercier *et al.*, 2013).

2.5.1 Comparison of the magnitudes of the evoked potentials

In general, in multisensory research, at the neural population level, there are several probabilities of multimodal interactions. They can either interact in a way of depressing or enhancing the signals of each other. These options are tested by comparing the magnitude of unisensory stimulus with the largest magnitude with the magnitude of the multisensory stimulus. As illustrated in the yellow, bottom panel in Fig. 9, if there is a depression, it can either be minimal, sub- and super-minimal when the smallest unisensory response is equal to, less than or more than multisensory responses, respectively. On the other hand, when the multi-sensory responses are bigger than the largest unisensory stimulus, it is an indicator of an additivity (green, top panel in Fig. 9). Sum of unisensory responses is compared to multisensory response, and the result is an indicator of the additivity, sub- or super-additivities.

Additivity analysis is a widely used method used to find out the interactions in multisensory integration. It is simply based on the comparison of the multisensory potentials and the linear summation of the unisensory potentials. Individual modality stimulus creates a neural pattern and the other modality of stimulus created another neural pattern in the neural oscillations. When the stimulus is given simultaneously, the same stimulus is expected to create the same patterns that should still be read in the EEG oscillations in a way that a linear summation of the individual oscillations, only if the neural patterns do not interact. But, in case of multisensory interaction, patterns interact at specific times and locations. In other words, ERPs induced by the multimodal $\boldsymbol{AV}$ stimulus is equal to the sum of ERPs evoked by unisensory $\boldsymbol{A}$ and $\boldsymbol{V}$ stimuli, if they are processed independently (Barth *et al.*, 1995; Stekelenburg & Vroomen, 2007). If bimodal response differs from the sum of two unimodal responses (supra- or sub-additive), this is attributed to interaction within the modalities in $\boldsymbol{AV}$ stimulus.

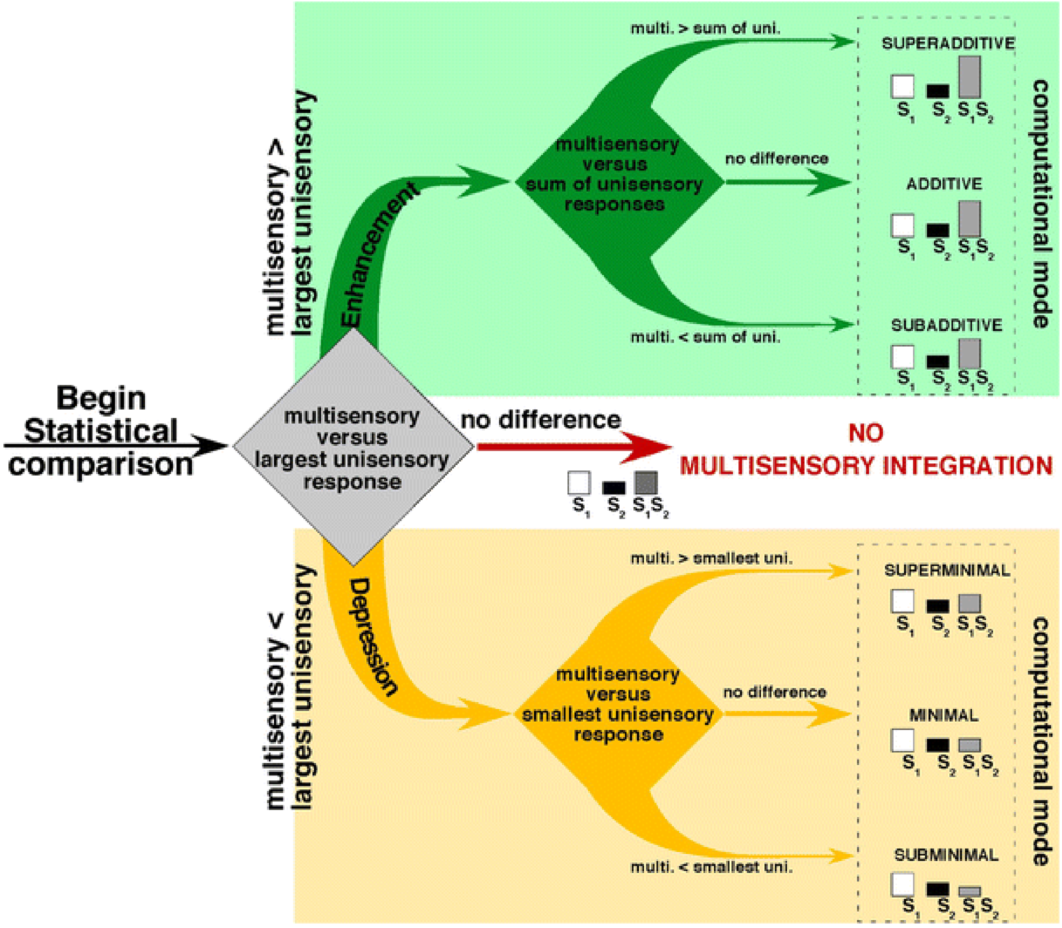


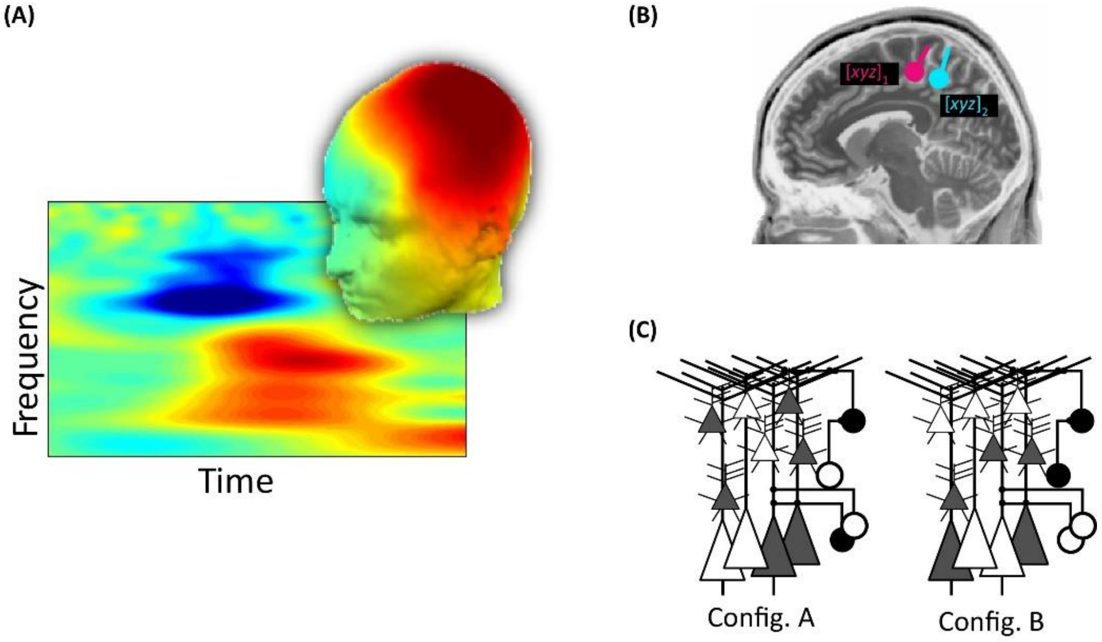
Fig. 9. A schematic of the standard methods for identifying and evaluating the computations underlying multisensory integration. First, the mean multisensory response magnitude is compared, using a statistical criterion, to the largest of the mean unisensory responses (gray diamond). If the multisensory response is significantly larger, multisensory enhancement is identified (green, top). If the multisensory response is significantly smaller, multisensory depression is identified (yellow, bottom). If multisensory enhancement is identified, the computation can be further subdivided into superadditive, additive, or subadditive enhancement by comparing the multisensory response to the predicted sum of the unisensory responses. Bar graphs provide examples of the response magnitudes evoked by crossmodal stimuli presented individually (S1, S2) and in combination (S1S2) for each computational mode. If multisensory depression is identified, the computation engaged can be further subdivided into super minimal, minimal, or sub-minimal depression by comparing the multisensory response to the smallest of the unisensory responses. Bar graph conventions are the same as above. (Adapted from Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009).

2.5.2 Oscillatory activities in crossmodal processing

EEG is a very fundamental imaging method in neuroscience that has very high time resolution, but low spatial resolution. EEG studies highlight the important role of ongoing signals in understanding the neuronal mechanisms that underlie cognitive processes. Different cognitive states and processes (e.g., memory, attention, perception, action, decision making, language processing) have been associated with specific oscillatory patterns of EEG signals (Buzsáki & Buzsaki, 2006). Every neuronal transmission creates electrical changes in the neuron membrane. But when a sufficient number of neurons are aligned and get sufficiently activated together, then that activation can be observed on the scalp by EEG in a specific form of an oscillatory pattern.

What you record in EEG is potential changes in time which are in an oscillatory pattern. These oscillatory patterns are mostly due to neuronal activities after removing the artifacts of muscle contractions, eye movement, heartbeats, and electrical noises from the environment. Oscillatory signals recorded in EEG can be decomposed into three components as illustrated in Fig. 10. EEG oscillations can be defined as a function of time, frequency, and location. In general, an oscillatory signal recorded in one EEG channel is considered as a composition of an infinite range of frequencies within varying amplitude and phases. However, in EEG, besides these components in one channel, due to the locations of channels, the topography of the signal on the scalp is also considered. Therefore, spatial information of the signal source can also be estimated. Eventually, simultaneous neuronal activities in the brain can be characterized by these components at various locations as a function of time. However, cognitive activities include more than one signal originating from various sources. Especially, when considering the multisensory environment, although the task is simple, multiple sources of information creates simultaneously traveling signals in the brain. One of the hardest problems is to estimate the travel of these several simultaneous signals in the brain. Even though EEG signals can be related to the cognitive tasks, they are a measurement of the actual neural activities. Therefore, it is important to keep the question in mind about what the EEG signals mean (for a discussion, see Cohen, 2017).

For example, a cognitive activity related to vision can be considered as the activity of occipital cortex, basically with the active involvement of other organs of the body. Considering the speed and the inter-connectivity of the nervous system, this activation can also be observed in other regions in a very short time lag. Even a simplest cognitive activity causes a chain of neural firings and transmissions distributed throughout the brain, which is apparently related to the other aspects of the cognitive activity. The neural mechanisms that operate in such a fast and decentralized form are difficult to understand by its own nature. Potential changes distributed over the scalp with EEG require extra attention.



Trends in Neurosciences

Fig. 10. Many features in EEG data are localized in time, frequency, and space, as in panel (A). At a given location, an oscillatory signal can be decomposed into its frequency feature, or for a given time and frequency window, the topography of the frequency power feature can be plotted. (B) Where do these features come from and what do they mean? Traditionally, this question is interpreted to indicate an XYZ coordinate in the brain that could have produced a given topography. (C) However, does one or the other solution which is shown here provide meaningful insights into how these EEG features arise from or are related to neural computations? Perhaps a more meaningful answer to the question ‘what does EEG mean?’ would come from determining the underlying microcircuit configurations that give rise to the features of the EEG landscape and that is consistently linked to cognitive processes. In other words, important advances in neuroscience will come from determining which functional/anatomical configurations in (C) could produce EEG features such as those illustrated in (A) (adapted from Cohen, 2017).

To make it clearer: a cognitive activity is a functional description or a serious of activities located in the brain and body. However, EEG is an electrophysiological record of the cortical activities. There are two difficulties at this point, (1) to identify the links between cognitive functions and cortical activities and (2) also to identify the link between EEG responses and neural activities. Therefore, attempts to put a direct link between EEG responses and cognitive activities might be misleading.

To grasp a better understanding of the distributed nature of neural responses, many methods have been developed in EEG analyses. Beyond a single channels oscillation, multichannel analyses are also done via one or more component of a channel from one region to components in the channels in different regions. For example, a high-frequency power component at one region may predict a phase component of a low-frequency oscillation at another region. With the various ways of analyses, EEG data provides us a very profound method of brain research.

There are many studies revealing the relationships between phase components in a specific region and the stimulus. Interestingly, in multimodal research, there are findings suggesting that phase resetting is a prominent mechanism that for interaction between early sensory cortices. Especially in case of audiovisual interactions, as earlier auditory stimulus excites the visual cortices, the phase of the oscillations is aligned to the onset of the auditory stimulus and the following visual responses interact with the aligned response (Mercier *et al.*, 2013; Escoffier *et al.*, 2015).

2.5.3 Phase-resetting framework for crossmodal interactions at early levels of visual processing

Phase resetting refers to the modulation of the phase of ongoing oscillations by external or internal event-related input. The crossmodal phase resetting hypothesis posits a shift of the phase of an ongoing oscillatory activity in one modality based on a sensory stimulus from another modality. Cortical excitability and hence sensory responses have been shown to be dependent on the phase of these oscillations. For example, when the auditory signal comes to the early auditory cortex, that signal resets the phase of the early visual cortex, which indicates an early interaction within these fields (Naue *et al.*, 2011; Thorne *et al.*, 2011). Another study testing the modulation of somatosensory input over auditory processing shows that somatosensory input can modulate the oscillatory phase of the oscillations in auditory cortex (Lakatos *et al.*, 2007).

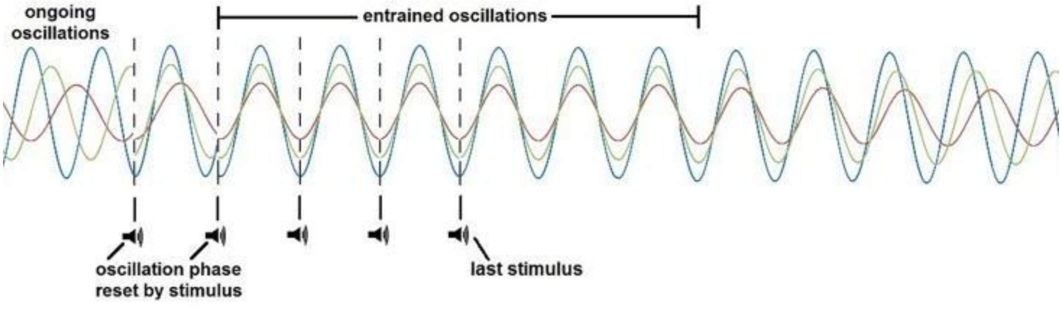


Fig. 11. An illustrating example of oscillatory entrainment by rhythmic auditory stimuli. The excitability of neural ensembles oscillates at various frequencies. The first stimulus resets the phases of ongoing oscillations. The second stimulus establishes a rhythm, and oscillatory frequencies adjust such that phases become aligned to this rhythm. Because of entrainment, neural ensembles are in a state of excitability when stimuli occur. Oscillations remain entrained for several cycles after the last stimulus before eventually “falling out of phase” due to frequency changes (Adapted from Lakatos et al., 2009).

A single attended or salient external stimulus can reset the phase of oscillations to a particular state of excitability within a sensory modality (Lakatos *et al.*, 2009). Stimuli in one sensory modality can also reset oscillatory phase in other modalities and this phase resetting can affect behavioral performance (Lakatos *et al.*, 2008; Fiebelkorn *et al.*, 2011; Romei *et al.*, 2012). For example, in one study visual stimuli caused phase reset in auditory cortex and oscillatory phase which in turn correlated with reaction time to auditory stimuli(Thorne *et al.*, 2011). Given that natural visual and auditory stimuli often co-occur (Thorne *et al.*, 2011), phase resetting by one type of stimulus may prepare a range of sensory areas optimally to process further incoming stimuli. In the case of rhythmic stimuli, repetitive phase reset is likely to underlie entrainment by initially aligning oscillatory phase and frequency to an external rhythm (Lakatos *et al.*, 2009) (Fig. 11). After entrainment is established, it can continue for some time in the absence of phase reset, as evidenced by entrained oscillations outlasting external stimulation(Mathewson *et al.*, 2012; Lakatos *et al.*, 2013). Like phase reset, entrainment can be crossmodal (Lakatos et al., 2008)

Experiments

The essential idea in the experiments is to pin down the neural mechanisms of the perceptual effect of a second modality (i.e. auditory, also pre-test visual in Experiment 4) in a percept of another primary modality (i.e. visual) by systematically varying the timing between the stimulations of these modalities. More specifically, in the experiments, neural underpinnings of time depended visual perceptions (i.e. relative timing, time interval or speed) are observed under varying SOAs of concurrent auditory presentations and in the last experiment aftereffects of both auditory and visual time intervals on visual apparent motion perception. Therefore, temporal ventriloquism in various perceptual contexts is investigated in both behavioral and neural activity. In all experiments, oscillations evoked by the multisensory conditions are compared to each other and they are compared to the unisensory conditions. Eventually, possible neural mechanisms are discussed with the support of the EEG data.

To test our hypotheses, four separate experiments focusing on the audiovisual interactions in time were designed (Fig. 12). In these experiments, the vision was considered as the primary modality while audition was secondary which provided additional temporal information. In the first experiment, only one flash is shown which leaves a mark of the single visual event in time, while in the second experiment, two event marks defining a time interval were presented. With a slight change, in the third experiment, the location of the consequent flashes has been changed. Change in a location within a very short time interval forms an apparent motion perception. The last experiment, with a slight change in design, investigated how being adapted to various time intervals influence the perception of apparent motion. To sum up, the first experiment focuses on the perceived time of an event, while the second one was about visually defined time interval perception, and speed perception in the third and fourth studies. As the secondary modality, pretest and concurrent auditory clicks had modulated the time onsets of the visual events. Eventually, this thesis aimed to investigate the neural mechanisms of these audiovisual temporal processes by collecting behavioral and EEG data in tandem

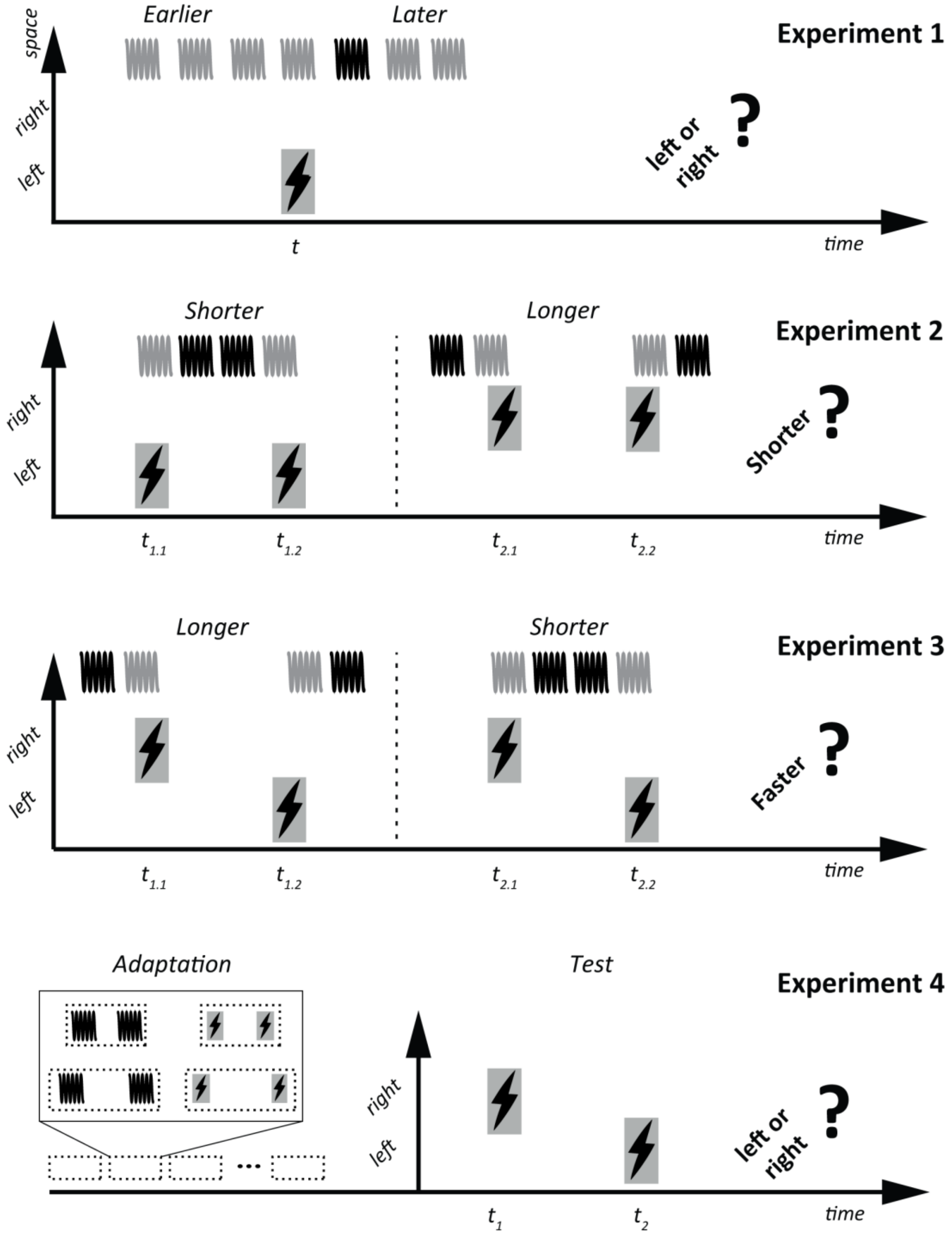


Fig. 12. Time and space diagrams of the visual stimuli and auditory stimuli in four separate studies. Flashes and clicks are depicted in the time-space diagram as flash sings and waves, respectively. Gray wave is for mild modulations, while black ones for extreme modulations in A, B, and C. (A) In Experiment 1, perceived time of the visual event which was presented at either on the left or right of the fixation point was modulated by auditory stimuli presented with varying SOAs. Participants were asked to report the location of the visual stimulus as early as possible. (B) Visual time interval perception was modulated by shorter and longer auditory intervals in Experiment 2. Participants compared two sequentially presented intervals at contralateral locations and reported the shorter-

perceived visual time interval. (C) In Experiment 3, speed perception of apparent motion was modulated by concurrent auditory presentations. Participants compared to two sequentially presented apparent motions and asked to report the faster-perceived one. (D) Finally, in a trial of the last experiment, participants are adapted to either short or long time intervals defined by the auditory or visual stimulus. Next, they asked to report the direction of consequently presented apparent motion.

CHAPTER 3

3 EXPERIMENT 1. AUDIOVISUAL INTERACTIONS IN THE TIMING OF A VISUAL EVENT

ABSTRACT

The integration of information from different senses is central to our perception of the world including the fundamental attributes of space and time. Audiovisual interactions have been particularly well studied in this context and various illusions have been developed to demonstrate strong influences of audiovisual interactions on our final perception (Chen & Vroomen, 2013). An interesting phenomenon is temporal ventriloquism, in which sounds drive the perceived timing of visual events. Even though temporal ventriloquism has been well studied at the perceptual level, the neural processes underlying auditory influences on visual timing is still unknown. In the current experiment, we investigated the neural correlates of temporal ventriloquism by collecting both behavioral and EEG data. During the experimental sessions, a brief visual bar was presented either at the right or left of a fixation circle. The visual stimulus was accompanied by an auditory click and the timing between two stimuli (-160 ms < SOA < 120 ms) was varied. The auditory click either led (SOA<0) or lagged (SOA>0) the visual bar in time. Observers (*N* = 20) were asked to report the location of the visual bar as soon as it was seen. The reaction times required to discriminate the location of the visual stimulus were significantly modulated by the SOA. Reaction times increased linearly as the SOA values were increased. In agreement with recent reports (McDonald *et al.*, 2013), the auditory click elicited an ERP over early visual cortex (occipital electrodes). Moreover, interactions between auditory and visual responses (superadditive effect) was observed on different sites varying by SOA, in agreement with the recent reports (e.g., Cecere *et al.*, 2017). Our results revealed that early occipital interactions were more salient when auditory signals precede the visual, however visual signals preceding auditory elicits mostly frontal super- and sub-additive nonlinear interactions. Interestingly, amount of additivity negatively correlated with the RT. On the other hand, the time-frequency analyses revealed that the crossmodal interaction led to an increase in 4-12 Hz (theta and alpha bands) power when the SOA value is between -80 and 80 ms. More importantly, for this range of SOA values, we also found changes in the temporal dynamics of 4-12 Hz power that paralleled reaction times to visual stimulus: the peak latency was increased linearly

as the SOA was increased. Together, our results, in combination with accumulating evidence (Naue *et al.*, 2011; Romei *et al.*, 2012), support an important role for low-frequency oscillations in crossmodal interactions at early levels of perceptual processing and different natures of the processing mechanisms of sound leading and light leading stimulation.

3.1 Introduction

Our cognitive system processes information coming from different channels of sensory modalities and integrates it smoothly. In most of the time, we are not even aware of that we combine the information in different sensory channels, rather all we are conscious of is some dimensions of the event itself such as its time and location. Having more senses provides more accuracy in the perception of such dimensions of the event. For example, one sense is less accurate about carrying time information of the event, while another is less accurate about estimating the location. Having multiple channels of information provides a better accuracy in perception. It was a long time ago that multi-sensory stimuli provided faster reaction compared to unisensory (Hershenson, 1962). The timing of an event in our perception is mostly driven by audition in such a way that, for instance, the rate of the auditory stimuli, flutters, affects the perceived rate of visual flickers more than flickers affect the perceived rate of flutters (Gebhard & Mowbray, 1959; Welch *et al.*, 1986). Likewise, an auditory click can change the timing of visual stimuli, which is known as temporal ventriloquism (Fendrich & Corballis, 2001; Morein-Zamir *et al.*, 2003; Recanzone, 2003). On the other hand, visual stimuli can affect the perception of the sound location (Howard & Templeton, 1966; Bertelson & Aschersleben, 1998; Alais & Burr, 2004a).

These phenomena were explained by the hypothesis that vision is specialized mostly for spatial information processing (location, tilt, or shape) while audition is for temporal (Welch & Warren, 1980), therefore each sense drives others in the dimension for which they are designed. Many recent studies have investigated neural mechanisms of audiovisual temporal integration. Specifically, auditory effects in the visual cortex have been shown by many EEG studies using different experimental designs (Busch *et al.*, 1989).

It was widely discussed that phase resetting is an important phenomenon in multimodal integration. As Naue et al. (2011) show, signals created by the auditory stimulus cause a phase resetting in the posterior areas. By the phase resetting mechanism, posterior regions are getting ready for the processing new neural signals. Multisensory data arrives at primary cortices closely in time and occurs in milliseconds short time intervals. The amplitude and phase of these oscillations have been shown to play an important role in crossmodal interactions. For instance, the influence of an auditory click on the sensitivity to a brief presented visual flash is reflected in amplitude and phase changes in EEG frequencies (10-30 Hz) over the visual cortex (Naue *et al.*, 2011).

As mentioned in general introduction, classical view claims that multisensory processing occurs in a hierarchical manner throughout convergence in the neural pathways. According to this view, the signal coming from different senses merge at higher association areas, since subcortical and early cortices are specialized mostly for unimodal processing (Meredith, 2002). The core idea of this approach is that signals from different senses are represented by the firing rates of the neurons. Group of neurons from one sense may have an excitatory or inhibitory effect on the others, which lead to having a crossmodal perception. Hence, quality of the perception is determined by the quantities of the different modalities (Senkowski *et al.*, 2008).

More recent studies, however, proposed that pure convergence model does not explain all aspects of cross-modal interaction (Ghazanfar & Schroeder, 2006; Kayser & Logothetis, 2007; Kayser *et al.*, 2008). Their views can be listed as follows. In the first place, it is claimed that we have strong evidence indicating that the crossmodal interactions occur already in the primary cortices, which hierarchical convergence view cannot account for. Second, convergence scenario cannot be as flexible as it can adapt to fast changes in multimodal data that creates completely new percepts. Third, convergence is a feed-forward mechanism where high-level representation does not have elementary properties, whereas the convergence model doesn’t explain how percept still has low-level information about the objects (Senkowski *et al.*, 2008).

The crossmodal phase resetting hypothesis posits a shift in the phase of an ongoing oscillatory activity in one modality, based on a sensory stimulus from another modality. Cortical excitability, and hence sensory responses, have been shown to be dependent on the phase of these oscillations. For example, when an auditory signal comes to the early auditory cortex, that signal resets the phase of the early visual cortex, which indicates an early interaction within these fields (Naue *et al.*, 2011; Thorne *et al.*, 2011). Another study testing the modulation of somatosensory input over auditory processing indicates that somatosensory input modulates the oscillatory phase of the oscillations in auditory cortex (Lakatos *et al.*, 2007).

Both behavioral and electrophysiological evidence suggests that auditory lead and auditory lag presentation of the stimuli are integrated into different neural networks (Cecere *et al.*, 2016, 2017). While auditory leading training effects the temporal simultaneity

Temporal ventriloquism has already been well studied at the behavioral level, however neural mechanisms of auditory influences on visual timing are still unclear. In the current experiment, we investigated the neural correlates of temporal ventriloquism by collecting behavioral and EEG data in tandem.

3.2 Materials and Methods

3.2.1 Participants

Twenty observers (7 females; mean age of 24.0 ± 3.7 years; 19-34 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 Ankara University.

3.2.2 Apparatus

We used MATLAB version 7.12 (The MathWorks, Natick, MA) with the Psychtoolbox 3.0 for stimulus presentation and data acquisition (Brainard, 1997; Pelli, 1997). Visual stimuli were presented on a 21-inch CRT monitor (1280 × 1024 pixel resolution and 100 Hz refresh rate) at a viewing distance of 57 cm. A SpectroCAL (Cambridge Research Systems, Rochester, Kent, UK) photometer was used for luminance calibration and gamma correction of the display. Sounds were presented via insert earphones (EARTone 3A, Etymotic Research, Village, IL) and amplitudes were measured by a sound-level meter (SL-4010, Lutron Electronics, Taipei, TW). The timing of auditory and visual stimuli was confirmed with a digital oscilloscope (Rigol DS 10204B, GmbH, Puchheim, Germany) connected to the computer soundcard and a photodiode (which detected visual stimulus onsets). All experiments were performed in a silent, dimly lit room.

3.2.3 Stimuli and procedure

The essential idea in this experiment is to pin down the role of the second modality (auditory) in a percept (of timing) of another primary modality (visual) by systematically varying a property (stimulus onset) of the secondary modality as a factor while keeping stimulus in the primary modality the same. As this experiment focuses on the perceived time of a visual event, only one flashed bar was shown as the primary stimulation. As for the secondary modality, physical time onsets of the clicks were modulated, as shown in Fig. 13.

A small red circle (0.3 deg diameter) at the center of the display served as a fixation target. In the time course of a trial, a visual bar (0.4 × 3.0 deg) was “flashed” (50 ms) either centered either to the left or right (1 deg), above (2.5 deg) the central fixation points as a visual event. The flashed bar was brighter (97 cd/m^2) than the gray background (20 cd/m^2). To keep the attention of the participant in the task, participants were requested to indicate the location of the bar (forced-choice task) whenever they see one, by a key press. They were informed that they will also be presented only sound in some trials where they were asked to do nothing. Reaction times of participants were an indicator of the timing of the perception, therefore participants were also asked to respond as fast as possible. Besides visual bar, a brief static “click” sound (20 ms) was introduced through headphones. Each click comprised of a rectangular windowed 480 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization and was binaurally introduced at 75 dB sound pressure level (SPL). Auditory stimulus accompanies visual stimuli (audiovisual: AV) in 8 different SOAs ranging from -160 to +120 ms by 40 ms steps (see Fig. 11). For evoked potential analyses, auditory and visual stimuli were also presented in modality-specific ways (visual-only: ***V***, auditory-only: ***A***). Participants were informed also about the ***A*** condition, where they were asked not to respond. A total number of 500 trials (50 trials in each condition) were presented in 5 blocks. Each condition occurred with equal probability within a block. Participants were encouraged to have a short break (approximately less than one minute) between blocks to maintain high concentration and prevent fatigue.

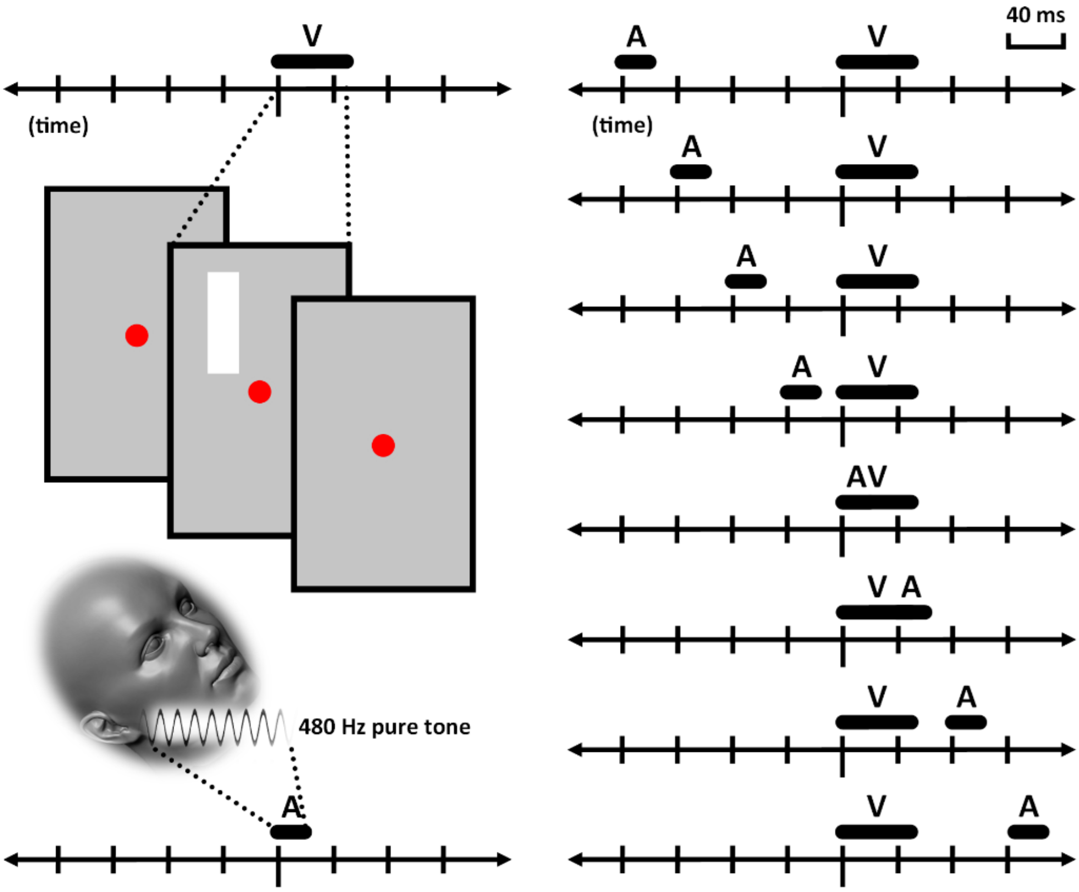


Fig. 13. Audiovisual experimental conditions used in single flash and single-click experiments displayed on the left. SOAs between visual and auditory stimuli are changed systematically by 40 ms steps. Sound either precedes or succeeds or accompanies synchronously with the flash bar in varying conditions. Two unimodal conditions: Display of visual stimulus on the spatial monitor screen and auditory stimulus is a 20 ms pure tone.

Trials started with 400 ms pre-stimulus interval and the flashed bar was shown either on the left or right pseudo-randomly. 800 ms post-stimulus interval followed the flash to wait for participants to react. Either reacted or not, experiment proceeds to the next trial after the post-stimulus interval. The inter-trial interval was randomly varied from 350 ms to 1050. Each click comprised of a rectangular windowed 480 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization and was binaurally introduced at 75 dB sound pressure level (SPL).

3.2.4 EEG data acquisition

Electroencephalography (EEG) data were recorded with a 64-channel MR-compatible system (Brain Products, GmbH, Gilching, Germany), using sintered Ag/AgCl passive electrodes mounted on an elastic cap (BrainCap MR, Brain Products, GmbH). The EEG caps included 63 scalp electrodes and two additional electrodes were placed, one electrocardiogram (ECG) electrode attached to the back of each participant to control for cardioballistic artifacts and one electrooculogram (EOG) electrode for recording eye activity. The placement of scalp electrodes was based on the international 10/20 system. Two of the scalp electrodes, FCz and AFz, was the reference and ground electrodes, respectively. A syringe and q-tips were used to apply conductive paste (ABRALYT 2000, FMS, Herrsching–Breitbrunn, Germany) and to reduce impedances in each EEG channel. Electrode impedances were kept below 20 kΩ (typically below 10 kΩ) and they were monitored throughout the experiment for reliable recording. EEG signals were digitized at a 5-kHz sampling rate, and band-pass-filtered between 0.016 and 250 Hz while raw data saving. BrainVision Recorder Software (Brain Products, GmbH) was used to store stimulus markers and EEG data on a secure hard disk for further analyses.

3.2.5 EEG preprocessing and analysis

We analyzed EEG data with BrainVision Analyzer (Brain Products, GmbH), the Fieldtrip toolbox (Oostenveld *et al.*, 2011) and our own MATLAB scripts (The MathWorks). For preprocessing, EEG signals were first downsampled to 500 Hz, and the cardioballistic artifacts were removed by the signal from ECG channel (Allen *et al.*, 1998). Then, the data were filtered through a zero-phase shift Butterworth high-pass filter (0.5 Hz, 24 dB/octave) and a 50-Hz notch filter (50 Hz ± 2.5 Hz, 16th order). Event maker has been set at the onset of the visual onset and this time point is considered as the point 0 in time. After filtering process, data were segmented into epochs from -600 ms (before the onset of flash) to 1000 ms (after the onset of flash). The infomax independent component analysis was applied on these epochs to remove common EEG artifacts (e.g., eye blinks). Also, each trial was screened automatically by artifact rejection criteria and manually by eye. Bad channels have been corrected by topographic interpolation (spline), on average In the automatic artifact rejection, any trial with oscillations over 50 μV/ms or a voltage change more than 200 μV was rejected. Trials with artifacts were rejected from the further analyses. After applying standard preprocessing procedure, on average 10.22% of trials were rejected per condition. For behavioral and EEG analyses, only trials which were correctly judged by participant were kept also in EEG preprocessing. Responses shorter than 200 ms and longer than 700 ms was considered incorrect. On average 4.4% of responses were incorrect.

After preprocessing, EEG signals from each specific electrode were averaged across trials to compute event-related potentials (ERPs) time-locked to the onset of the visual stimulus. A low-pass filter (40 Hz cut-off frequency) was applied to further smooth the ERPs. Baselines were also computed from a -100- to 0-ms time interval prior to the onset of the first stimulus and they were subtracted from the ERPs of each condition. As we examined the effect of auditory stimuli on the visual perception, auditory-visual linear summation (A+V) was obtained by adding auditory responses (A) to visual responses linearly (V). **A** responses were shifted in time accordingly in each corresponding SOA condition, which can be notated as follows: $\mathrm{AV}_{\mathrm{SOA_i}}\ ?=\ (\mathrm{A}+\mathrm{V})_{\mathrm{SOA_i}};\ (\mathrm{A}+\mathrm{V})_{\mathrm{SOA_i}} = \left(\mathrm{A}_{\mathrm{SOA_i}}\right) + \mathrm{V}$, where $\mathrm{SOA} = -120{:}\,40{:}\,160$. After the summations, ERPs are reepoched to -400 to 600 ms and baseline corrected again before statistical running tests. To determine specific components and electrodes significantly affected by SOAs, ERPs were compared between AV and A+Vs across 8 AV conditions using pointwise running *t*-tests (paired samples two-tailed). Furthermore, we introduced basic corrective steps to prevent false discoveries. A significant difference in conditions was defined when it was stable for at least 20 ms of contiguous data (i.e. 10 consecutive points at a sample rate of 500 Hz) and present on at least three neighboring electrodes meeting 0.05 alpha criterion (Guthrie & Buchwald, 1991; Picton *et al.*, 2000; Senkowski *et al.*, 2007; Vroomen & Stekelenburg, 2010; Keil *et al.*, 2017). Over the identified time windows and scalp sites based on the outcome of pointwise running *t*-test and corrective steps, we estimated the average ERP magnitudes for each participant.

3.2.6 Time-frequency and phase analyses

Oscillatory activity was transformed into time-frequency domain by morlet-wavelet transform (see Naue *et al.*, 2011). As the basic assumption suggests, oscillatory responses may occur either with a variability in phase (induced response) or strictly phase locked to the stimulus onset (evoked-response) (Başar-Eroglu *et al.*, 1996). Therefore, we used different methods of evaluation. We analyze the responses phase locked to the stimulus by the wavelet transforms applied to the averaged ERP. However, to reveal the non-phase-locked portion of the response, each trial is first transformed to the frequency domain and then absolute values of the resulting wavelet transforms are averaged. Therefore, this measure represents the *total activity*, comprising the phase-locked and non-phase-locked part of the response. The wavelet analysis was performed for 1 Hz frequency bins in the frequency range of 1–80 Hz. Phase locking to the stimulus across trials was quantified by a time-frequency version of the so-called mean resultant length (Fisher, 1995) as has been described before (Tallon-Baudry *et al.*, 1996; e.g. Fründ *et al.*, 2007). The *phase-locking factor* (PLF) yields values between 0 and 1, where 1 indicates perfect phase locking across trials, while 0 indicates a constellation in which the phases exactly cancel each other out, as is the case for a uniform distribution of phases across trials. The time-frequency planes were transformed to a decibel scale for evoked and total responses if absolute values were presented, indicating the change relative to a baseline that extended from 400 to 200 ms before the onset of the stimulus. Phase locking to the stimulus was related to the same baseline by subtracting the average phase locking of this window. For statistical analysis, the resulting time-frequency representations were pooled into a region of interest (ROI) and averaged across those electrodes that exhibited the strongest response to visual stimulation. The mean amplitude and phase-locking values were calculated for three analysis windows in three different frequency bands at posterior electrodes, averaged across all participants. Wavelet parameters are optimized for a time-frequency resolution to prevent from leaking between frequencies due to wavelet transform and then baseline correction is performed on the resulting transformations.

For statistical analysis, the resulting time-frequency representations were pooled into two regions of interests (ROIs) located over posterior (PO3, POz, PO4, and Pz) and frontal (AF7, AF3, F7) regions that exhibited the strongest response to visual stimulation. The mean amplitude and phase-locking values were calculated for various frequency bands at both posterior and frontal electrodes, averaged across all participants.

According to the evoked V responses, the time-frequency windows for statistical analysis were defined as follows: low beta-band activity: 12–20 Hz; alpha-band activity: 8–12 Hz; and low and high theta-band activities: 4–6 Hz, 6–8 Hz. Frequency bands were averaged in frequency dimension over all time points. Then, amplitudes and latencies of peak values were estimated from 2D plots of 4 frequency band for 20 participants, 4 modalities ($\boldsymbol{AV}$, $\boldsymbol{A}$, $\boldsymbol{V}$, $\boldsymbol{A}+\boldsymbol{V}$) and 8 conditions (various SOAs). One-way ANOVA have been applied on amplitudes and latencies to test the effects of various SOAs over 4 modalities.

3.2.7 Behavioral data analysis

For behavioral data, we only used trials included in the EEG analyses. In other words, the trials excluded during the EEG preprocessing stage were not considered for the analysis of the behavioral data. We calculated average performance across participants for each stimulus condition. To determine whether the effects of SOA significant, we applied one-way repeated-measures ANOVA to this behavioral measure. A simple linear regression was calculated also to predict RT based on SOA.

3.3 Results

In the present experiment, we used high-density EEG to investigate the neural mechanisms of auditory influences on perception time of a flash. We recorded ERPs during a classic two forced-choice tasks in several conditions (AV pairs with 8 SOAs) and used additivity analysis and analyses on power and phase components of time-frequency decompositions. Under the hypothesis that different neural networks and mechanisms are involved in audiovisual temporal binding depending on SOA, quantitatively different spatiotemporal interactions patterns are expected to emerge across different SOAs.

3.3.1 Behavioral results

As Fig. 14 shows that participants responded to the visual stimulus with a mean reaction time (RT) of 327 ± 22 ms. RTs of 20 participants significantly increased by early auditory signal condition to later in a linear fashion ($F7,133) = 61.08$; $P < 0.001$), in agreement with the findings in Fendrich & Corballis (2001). In addition, RTs for the visual-only condition was bigger than the synchrony condition. A significant regression equation was found, with an R^2 of 0.95. Predicted RT is equal to 327.4 + 0.2 * SOA, where SOA is coded or measured as ms., and RT is measured as ms. Participant's RT increased .2 ms for each ms. of SOA. SOA was a significant predictor of RT. Additionally, in line with Diederich et al. (2012), an oscillatory pattern overlapping the individual participants' increasing trend as a function of SOA can be observed in light gray plots of the Fig. 14. Relationship of the reaction times and alpha oscillations were also shown before (Callaway & Yeager, 1960; Dustman & Beck, 1965).

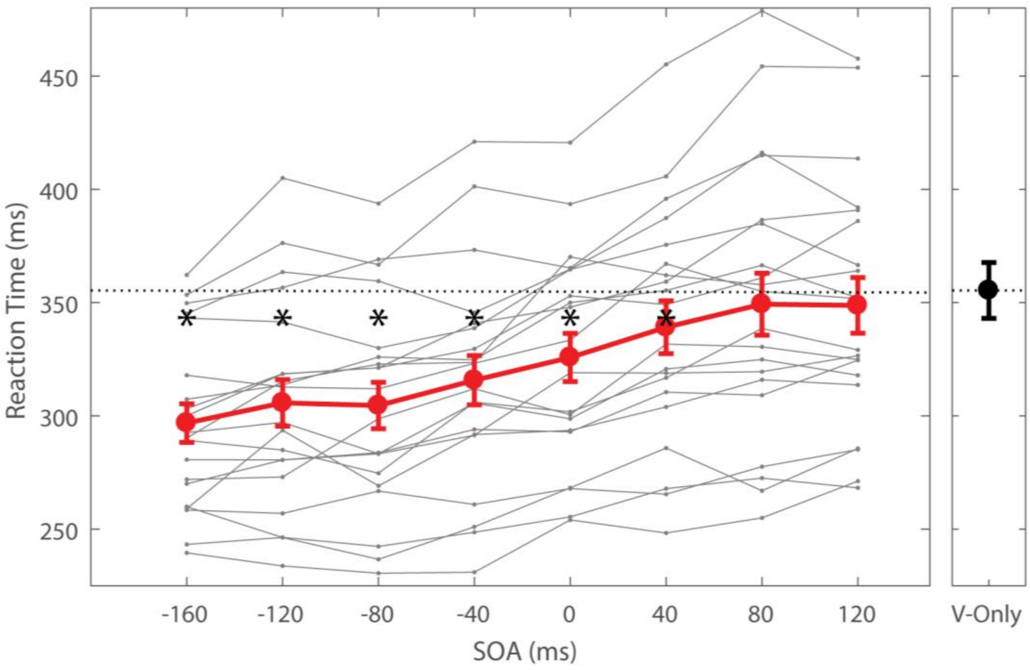


Fig. 14. Display of reaction times (RTs) to a visual stimulus according to different auditory stimulus timing. A significant linear increase in RTs for different SOAs was observed, especially between -80 and +80. For conditions SOA ≤ 40 ms, participants (n = 20) responded significantly earlier than visual-only condition ($P < 0.05$). Light grey plots are for RTs of each participant. Red plot and error bars correspond to averages and ± SEM, respectively. The dashed lines and the graph on the right with the black error bar show the mean RT for visual-only condition and ± SEM.

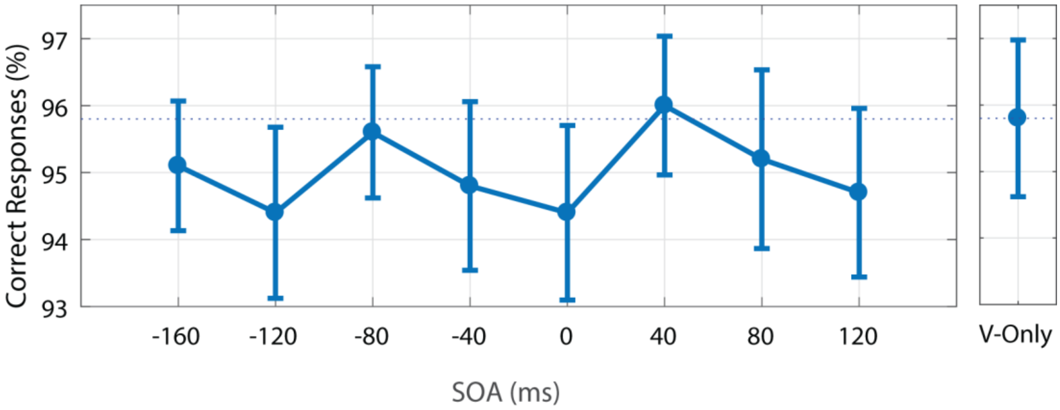


Fig. 15. The mean percentage of correct response values averaged across participants (N = 18). Error bars correspond to ± SEM. The dotted line indicates the mean value for the no adaptation condition, and the error bar placed over the symbol at the end of this line represents ± SEM.

As expected, all observers reported the location of the flashed bar with more than 90 % accuracy for all conditions (see Fig. 15). This indicates that all participants were paying attention and executing the task correctly during EEG recording. There was no effect of SOAs on the percentages of correct responses and none of the SOA conditions was significantly different than the visual-only condition.

3.3.2 Uni- and bi-modal responses and nonlinear interactions observed at various regions

EEG analyses have been carried out to find out how these perceptual effects of auditory timing overlap with the nerve responses in the different regions of the cortex. Fig. 16 shows the average ERPs for synchronous bimodal (**AV**), auditory-only (**A**) and visual-only (**V**) conditions from occipital, temporal, and frontal ROIs. The visual-only stimulus causes significant activation (diverge from baseline) in the area above the early visual cortex, and these occipito-parietal locations are similarly activated also for auditory-only stimulus, but with a smaller amplitude. A reverse pattern can also be observed in temporal channels which are above the early auditory cortex. In V condition, as expected, P1, N1, and P2 were observed clearly at 120, 180 and at around 290 ms respectively. From the topological plots, P1 was mainly at around V1 area and N1 was mainly at later visual cortices, and probable motor activity at central area accompanies with the occipital cortex at P2. As for **A** condition, activity travels from primary visual to the later areas.

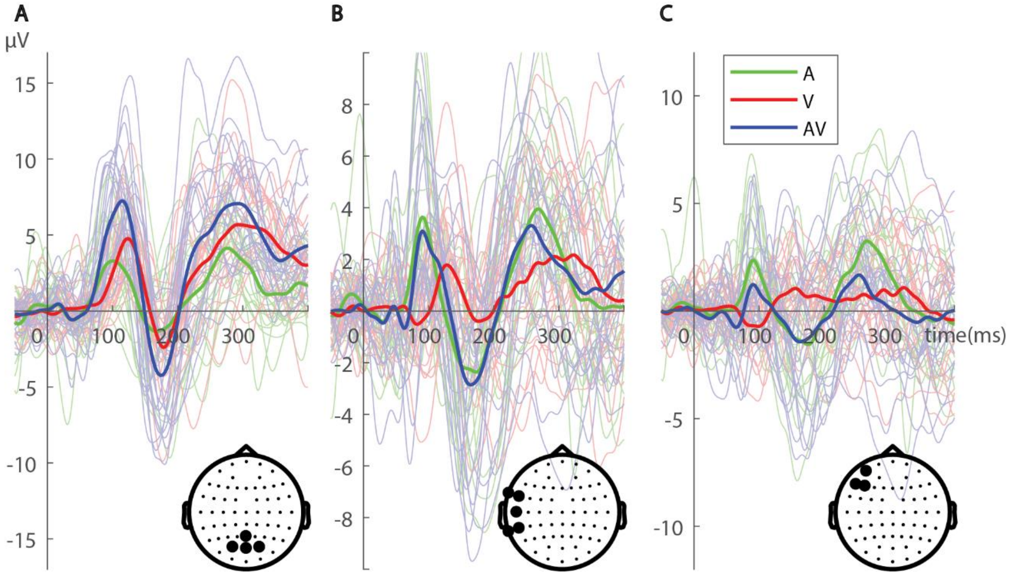


Fig. 16. EEG responses (ERP) against concurrent audiovisual (**AV** SOA = 0) and uni-modal stimuli (**V** and **A**). The blue line shows ERPs that are audiovisual (**AV**), red visual-only (**V**) and green auditory-only (**A**) conditions. Physical stimuli are presented at 0 ms. (A) The amplitude of the activity for the **AV** condition in occipito-parietal is expected to be higher than that of the single modal conditions as expected. Although the responses with only audiovisual conditions are only visually low, the reference is significantly larger than the active. ERPs induced to these experimental conditions in (B) temporal channels and (C) frontal channels.

3.3.3 Two distinct ERP interaction patterns observed as a function of audiovisual SOAs

As we wanted to determine how auditory stimulus affects the neural responses in various conditions, we should observe significant interactions in comparisons of between **AV** and **A**+**V** ERPs for different SOA conditions. Moreover, we reasoned that if there are specific times and locations for the audiovisual interactions, in all conditions, sub- and super-additivities should be observed at those time windows (TWs) and locations. Otherwise, we should observe some spatiotemporal shifts in interactions as a function of SOAs. To test these possibilities, we performed pairwise comparisons (pointwise running *t*-tests with basic corrective steps) between **AV** and **A**+**V** ERPs across all individual electrode locations for each SOA, separately. For all conditions, the pairwise comparisons revealed significant differences between **AV** and **A**+**V** responses in all conditions. However, spatiotemporal patterns of interactions vary as a function of SOA between modalities. We identified three different TWs in which the audiovisual interactions take place and two groups of channels: I: 60-100 ms, which fits on the rising edge of the P1 peak; II: 130-170 ms, rising edge of N1; and III: 200-250 rising edge of P3. After the third TW, the interactions are so close to the reactions that they may not influence the reaction. As shown in Fig. 17A, early super- and sub-additivities are observed in the TW-I and II, respectively over the posterior channels. However audiovisual interactions shift to the frontal channels for those conditions that the sound comes in synchrony or after the flash. Interactions were observed mostly at TW-I and III.

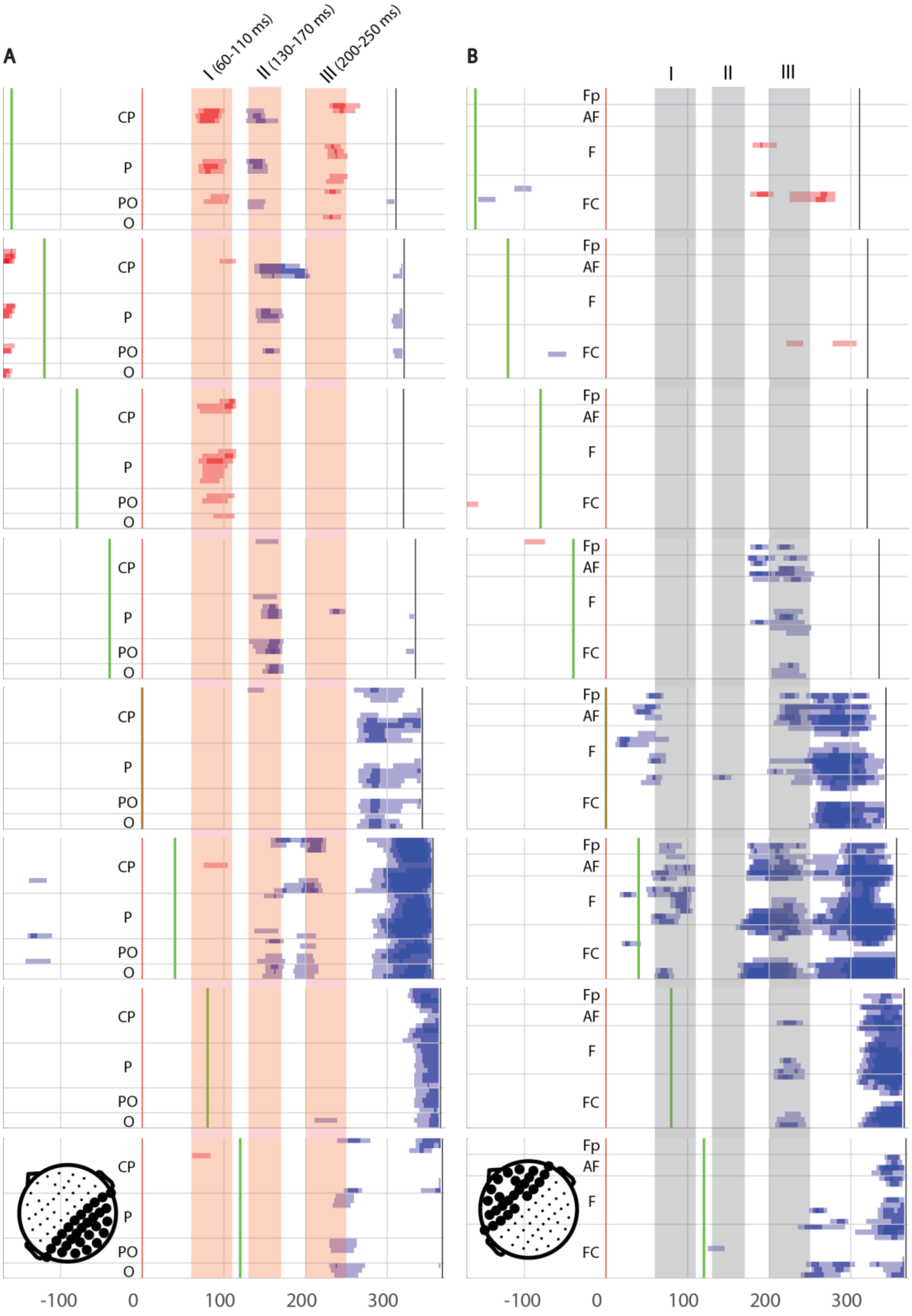


Fig. 17. Spatiotemporal profiles of audiovisual interactions in the single click single flash experiment. AV and **A+V** responses were compared with the help of the point-wise running *t*-test (two-tailed, *P* <0,05) whether the super- or sub-additivity was significant or not and the significant differences were

represented in the colors of the blue and red where *P* values were significant. The horizontal axis indicates the time and the vertical axis indicates the electrode location. The electrodes are shown in four separate horizontal rows.

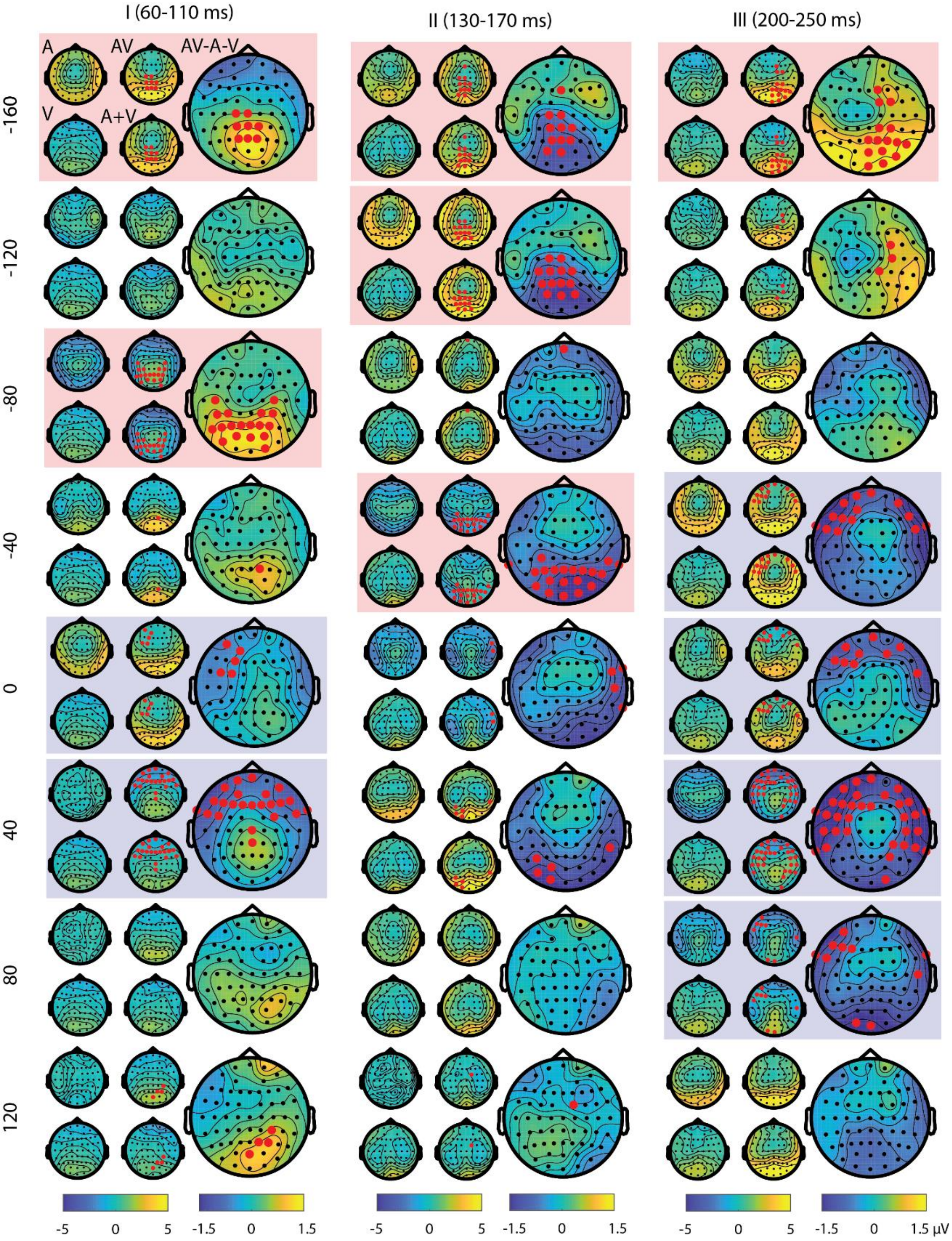


Fig. 18. Topographies of audiovisual interactions in various time windows. For 8 **AV** SOAs in rows, topographies of mean responses for 3 TWs (I:60-110, II:130-170, II:200-250 ms, in columns) have been shown. For each SOA/TW cell, synthetic A, V, AV, and synthetic A+V response distributions were given in small topographies, and AV-(A+V) responses are shown in the big topographies. Red

dots indicate the channels where the comparison was significant ($P < 0.05$) in the given time window. Shades behind the topographies indicate the cells where significant interactions were observed, red shades stand for the cells where the interaction is mostly observed at posterior channels, while blue is the ones at frontal channels.

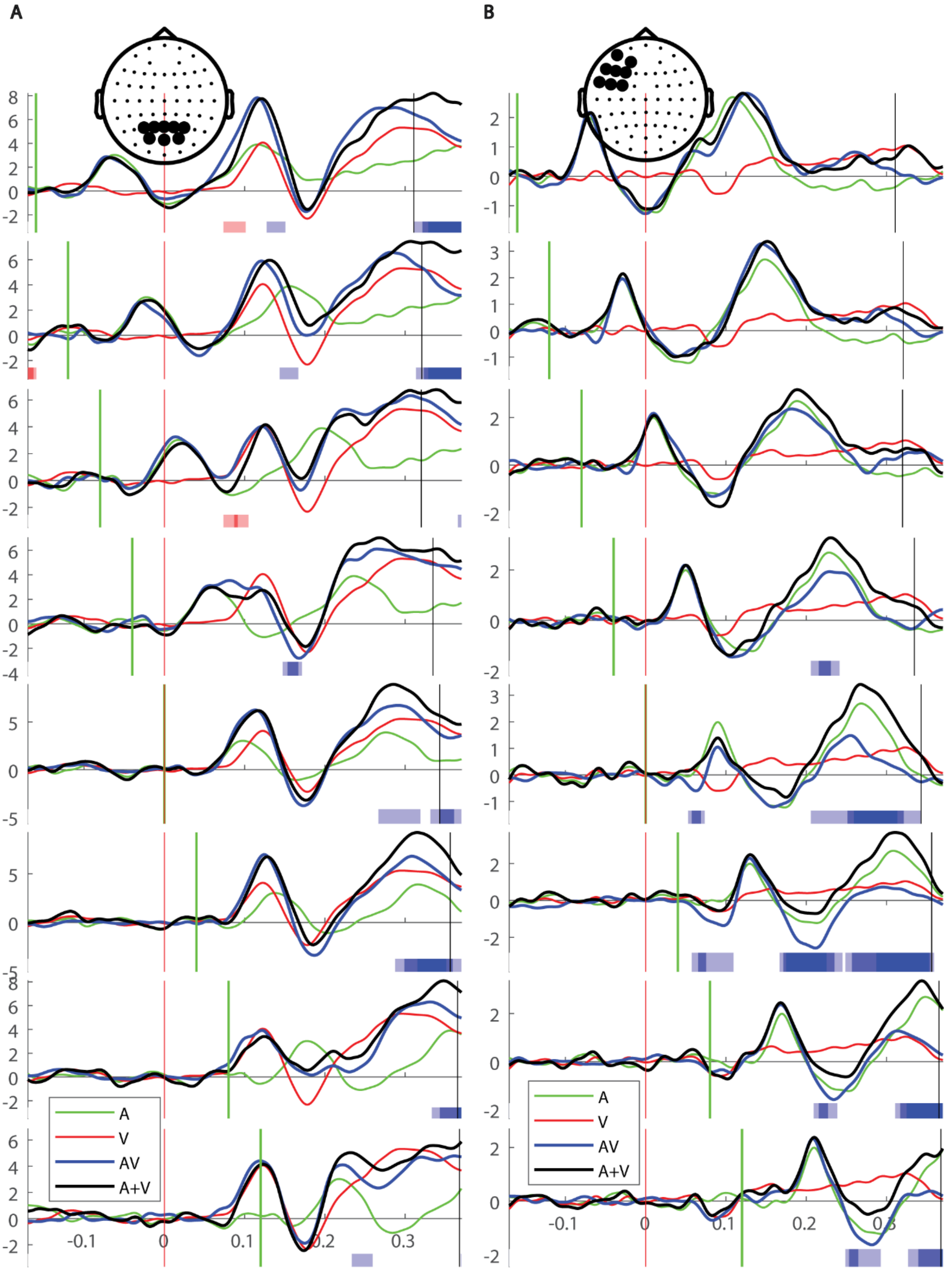


Fig. 19. Average ERPs from two different ROI centers (A: occipito-parietal, B: frontal). Each auditory timing condition (SOA) is given in different rows. The ERPs elicited by audiovisual (blue:

AV), visual (red: V), auditory (green: A) and the sum of auditory and visual (black: A+V) stimulus. The red shades on the time axis show the intervals where the difference between AV and A+V is significant (P <0,05). Red and green vertical dashed lines show the physical start time of auditory and visual stimuli. Black vertical dashed corresponds to the average reaction time to visual stimuli.

In Fig. 18, topographies confirm that when the sound comes before the flash ($SOA \leq -40$), interaction is observed mostly at *parietal* channels at early time windows (TW-I and II). However, when the sound comes in synchrony or just after the flash ($0 \leq SOA \leq 40$), early (TW-I) and late interactions (TW-III) were observed at *frontal* channels. However common channels of interactions have been selected for further investigation of the ERPs.

Average ERPs from common channels where an interaction observed at occipital and frontal regions reveals that early sound accelerates either first P1 or second N2 peaks, which corresponds to the TW-I and TW2, respectively. However, when sound presented closer to the flash in time, frontal interactions get foreground, instead.

To sum up, three main effects can be observed from all conditions. First, super- and sub-additive interactions are observed in the occipito-parietal channels (Fig. 19A: Occipital and Parietal Areas) during the rising times of the first and second hills (60-110 ms and 130-170 ms intervals). This super-additivity effect in first and sub-additivities in the second hill are salient especially for the conditions where the click comes earlier then flash (SOA ≤ 0). This means that clicks that come just before the flash creates a higher amplitude in occipital channels than a linear addition of plain auditory and visual components. In these conditions, this additive effect can also be interpreted as peak shifts for AVs to earlier time compare to A+V, therefore these sub- and super-additivities can be interpreted because of the peak shift. This might be an indicator of the fast responses for the first 4 conditions. Second, at a later time (250+ ms), again in the occipital area, when the sound comes closer in time (SOA 0, +40), we observed another significant sub-additive effect, however, it is so close to the reaction that might not be related to some other task-relevant interaction. Third, interestingly, when click comes closer or later than the flash onset, nonlinearity can be observed in frontal channels. This sub-additivity is observed starting 200 ms before the reaction.

To have a better understanding of how additivities change in time over various SOAs, additivity values for all conditions were combined in Fig. 20. In terms of the interactions for both occipital and frontal channels, there is a similar pattern in time. But interestingly, for smaller SOAs, time intervals corresponding to the first and second peaks show stronger early super- and sub-additivities respectively in occipito-temporal regions (Fig. 19). But frontal channels do not have an interaction for small SOAs (see Fig. 19B, panels in the first three rows). As for the SOAs getting bigger, the role of the occipital decreases but a stronger role of frontal channels can be observed. It is important to note that although the interaction pattern looks similar, differences are all about the small quantitative changes.

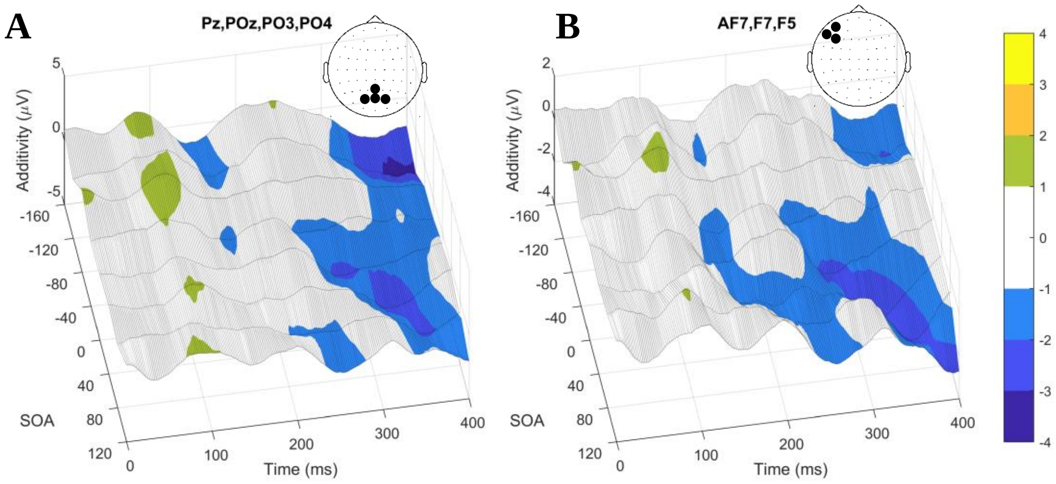


Fig. 20. Additivity distributions in time and various SOAs. Z-axis represents the additivity value which is calculated by subtracting the A+V from AV. Y-axis represents the SOAs while X-axis depicts time. (A**)** The left panel is for additivities of the occipital ROI. (B**)** The right panel is for frontal ROI.

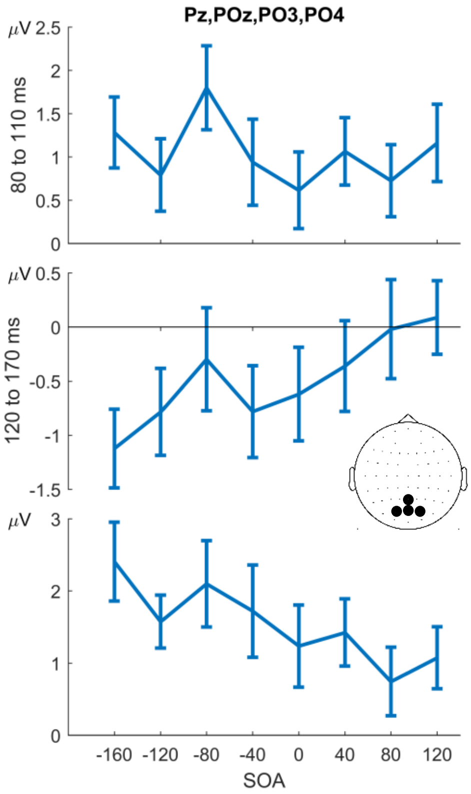


Fig. 21. Amount of additivity for P1 and N1 peaks. First two panels are for the time intervals of the first and second peaks while the third panel represents the difference between the first two panels. This value may represent a total amount of additivity, which corresponds to the time shifts in P1 and N2.

More detailed analyzes were performed for the time interval corresponding to the first rising peak and second rising peak time intervals. Fig. 20 shows the amount of additivity of the second hill, which is negative, is subtracted from the amount of additivity at the first hill, to determine the sum of the shifts in the early hills. The result shows the significant SOA effect on the shift of first two peaks. It is noteworthy that this result is of a nature that can predict the reaction times to the visual event.

3.3.4 Time-frequency analyses

To understand the frequency ranges of audiovisual interactions, wavelet transform analysis was performed over two channel groups analyzed above. For the first part of the analyses evoked responses have been obtained. As shown in Fig. 22, the evoked responses for visual (V), auditory (A) and audiovisual conditions are mostly strong at theta (4-8 Hz), alpha (8-12 Hz) and beta (12-20 Hz) frequency bands. Additionally, oscillatory responses were observed also at the higher frequency gamma band (40-60 Hz). Analyses show that the theta band is divided into two separate frequency bands (4-6, 6-8 Hz) because two different activity clusters are seen in the theta frequency band in wavelet transforms.

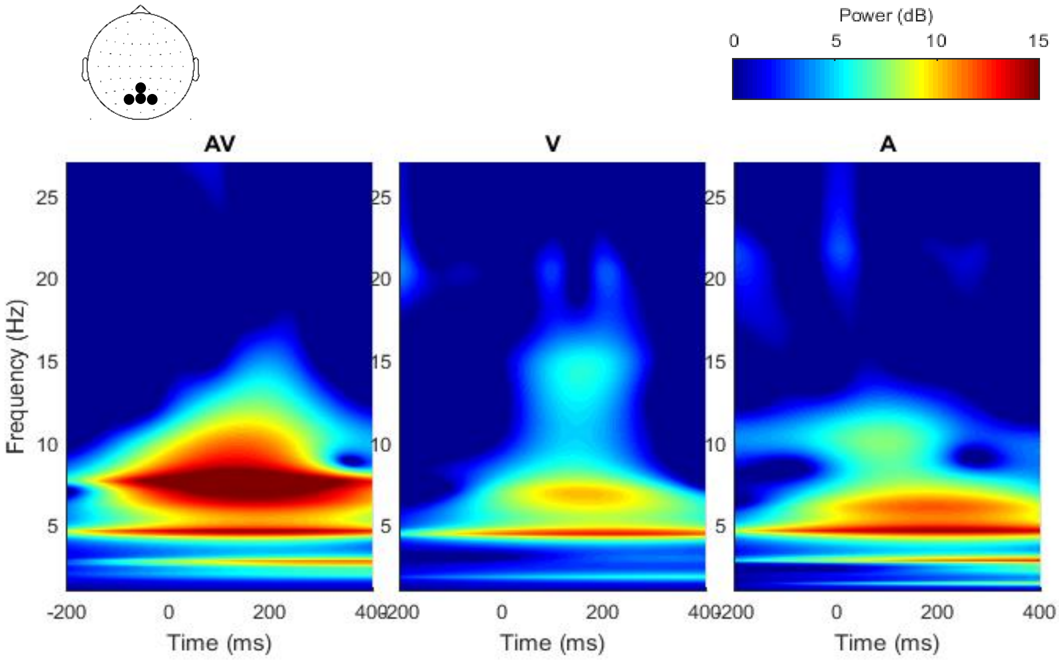


Fig. 22. Wavelet transformations of the occipito-parietal part of audiovisual (AV SOA = 0), visual only (V) and auditory only (A) conditions. Transforms were obtained with mean signals over all participants.

First, as an indicator of the timing and power of the neural responses, peak analyses have been applied on the ERPs that are averaged over all participants. According to both unimodal and multimodal conditions, some frequencies of interest have been defined, which are very close to the frequencies given above, and peaks were detected on the 2D surfaces of the responses.

3.3.5 Frequency bands predict the visual and auditory stimulus timing and interactions for AV evoked potentials

In the posterior electrodes as shown in Fig. 23. A and B, peaks latencies on the high-frequency band (40-60 Hz) are observed just after the auditory stimulus (~40 ms), which can be observed also at A-Only condition (Fig. 23, fourth column). Therefore, evoked activity for the auditory stimulus can be observed at the 40-60 Hz frequency at posterior electrodes. We observed the similar pattern at the frontal channels, too (see Fig. 23C).

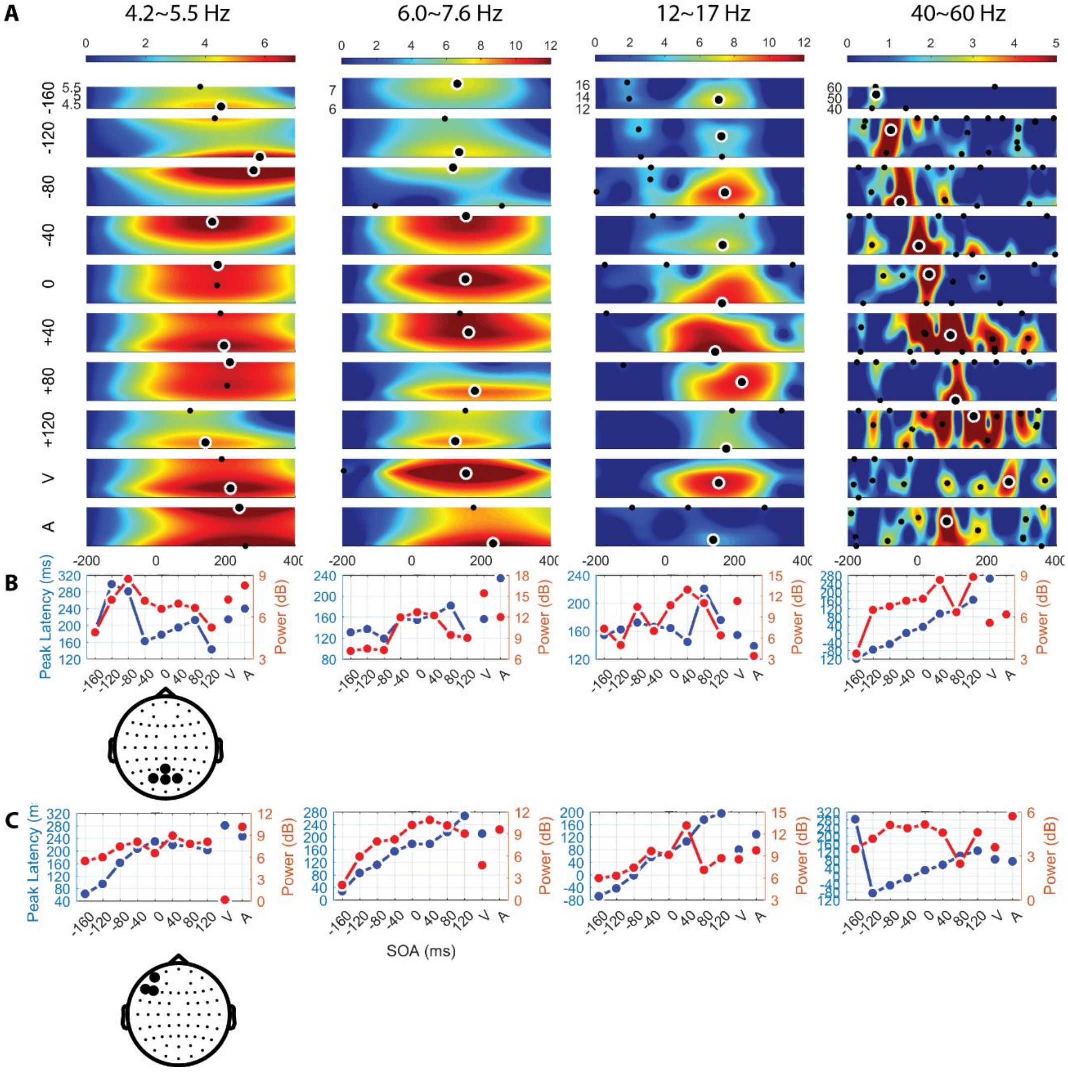


Fig. 23. (A) Time-frequency plots of evoked activity from the posterior ROI, averaged across all participants (N=20) within four frequency ranges (lower frequency to higher depicted in panels from left to right, respectively). Panels from top to bottom represent time-frequency maps for the bimodal (AV) conditions, as well as unimodal visual (V) and auditory (A) conditions. Black dots illustrate

the peak points in the given range and the one with a white circle is the one with the biggest amplitude values given next to it. (B) Peak latencies (blue) and powers (red) of the evoked activities marked in Fig. 23A for various SOAs of bimodal AVs, and unimodal V, and A conditions in various frequency ranges (shown distributed in columns). (C) Peak latencies and powers for the left anterior channels in four frequency ranges.

As for high alpha band of 12-17 Hz (Fig. 23, third column), a clear peak was observed at almost every condition that has visual stimulus, at 160 ms after V stimulus. For those SOA ≤ 0 ms conditions, peak time is stable at 160 ms, which may be interpreted in the way that this frequency activity is related to the visual stimulus directly. Probably, for those conditions of SOA ≥ 40, as the auditory evoked potentials entrain the visual stimulus at this frequency band. As for the power changes, an interesting oscillatory pattern was also observed for SOA ≤ 0 conditions.

For the lower frequencies, powers are very strong at those conditions of timings of the peaks shift slightly with SOAs. Especially for conditions in the middle (-40 ≤ SOA ≤ 80), there is a strong amplitude increase compare to other conditions. Latencies for these frequency peaks are in an increasing trend, which might be a strong indicator of the multimodal interaction. Qualitative change in both frequency and amplitudes of the evoked powers at -80 conditions, which looks like a threshold value for the oscillatory power changes.

Evoked potentials at the anterior channels have followed the auditory stimulus in the gamma band (40-60 Hz), very similar to posterior channels, at 40 ms after the auditory stimulus. Interestingly, evoked potentials followed the auditory stimulus both in a high alpha band with ~40 ms delay and in high gamma band with ~80 ms delay. Furthermore, latencies of theta activities on two distinct frequencies have also shown an interesting pattern. For this frequency ranges, there is no direct fixed relationship with any of the unimodality. For low theta (4.2~5.5 Hz) visual only stimulus not even evoke a potential but auditory signal evokes a potential which shift with the stimulus onset till SOA=-40 and for later conditions latency get stable, which must be due to the visual interaction.

To investigate the effect of auditory timing on these determined frequency ranges, the *peak latency* and the *average amplitude* around the peak amplitude at each of these intervals were calculated also from each participant. These further analyses had been done also for AV-A evoked potentials. The assumption behind this analysis was that the changes in the average amplitude showed the changes in the power of the oscillations more, while the changes in the delay tend to reflect the changes in the dynamics of the neural responses.

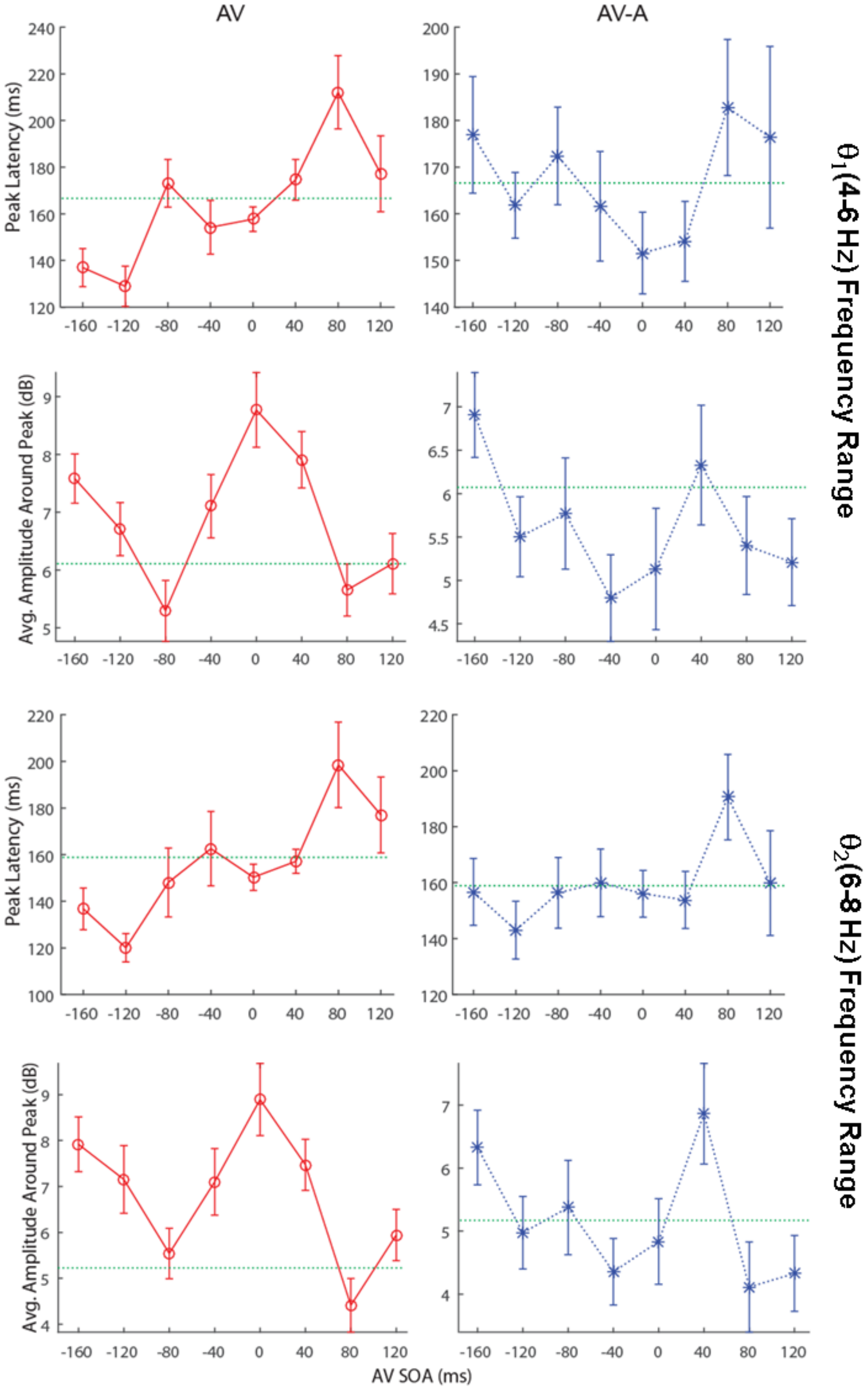


Fig. 24. Peak analysis over high and low theta bands (6-8 Hz) for AV conditions (red line) and V-only (green horizontal line). (A) Peak amplitudes change with the change of SOA values in an

oscillatory fashion. (B) Peak latencies increase with increasing SOA values, which correlates with RTs.

In Fig. 24, the obtained latencies and average amplitude values for the theta frequency band are given. Significant changes have been observed both in multisensory audiovisual (AV) and calculated audiovisual interaction values (AV-A), which were obtained by subtracting auditory component from audiovisual. Auditory timing effect is salient in these frequencies, especially in the 6-8 Hz range. In this theta frequency range, AV peak timing varies significantly with SOA ($F (7,144) = 3.29$, $P < 0.01$) and the delay time increases as the SOA increases. When three extreme SOA points were excluded in audiovisual interaction (AV-A) values, a significant and similar change due to auditory timing is observed ($F (4,90) = 2.26$, $P < 0.05$). In addition, the mean amplitude values for the two cases vary significantly depending on the SOA (AV: $F (7,144) = 4.95$, $P < 0.01$; AV-A: $F (7,144) = 2.26$, $P < 0.05$).

In both cases, the amplitude values increase and decrease in a fluctuating manner. The phase coherence analysis performed (in other words, the phase consistency index, intertrial coherence values) reveals that these changes are a significant effect of phase-resetting mechanisms Fig. 25. In accordance with the evoked potentials, a clear phase synchrony was observed in high frequencies, with very similar latencies. For lower frequency bands, phase synchrony gets its peak at around 160 ms after visual onset and the coherence level fluctuates. Period of this fluctuation is 160 ms, therefore its frequency is at around 6 Hz. This result suggests that auditory signal causes an oscillatory effect which travels to posterior channels. And very interestingly, the peak of the synchrony was observed only at around 160 ms after the visual onset. This is a good indicator of the idea that the phase synchrony is due to the visual stimulus, while SOA of the auditory stimulus modulates the level of the phase coherence.

3.3.6 Intertrial phase coherences were in accordance with the total evoked potentials

With the assumption that ITC is a prominent mechanism for inter-cortices interactions, Phase coherence analyses were done at the occipital channels. We expected that auditory responses should modulate the phase component of the occipital channels in the early time windows. In line with the previous studies, we expected to see the components that are corresponding to the auditory and visual responses and also for bi-sensory responses (Naue *et al.*, 2011). ITC (Fig. 25) in gamma (40-80 Hz) band reveals auditory component following the auditory stimulus with a 40 ms latency. Furthermore, latencies of theta frequency phase coherences relate to the multimodal interaction. ITC levels also change as a function of SOA. These results are very similar to the peak latency and average amplitudes around the peaks. This similarity indicates that the oscillatory power increases in the relevant frequency bands are due to the phase coherence.

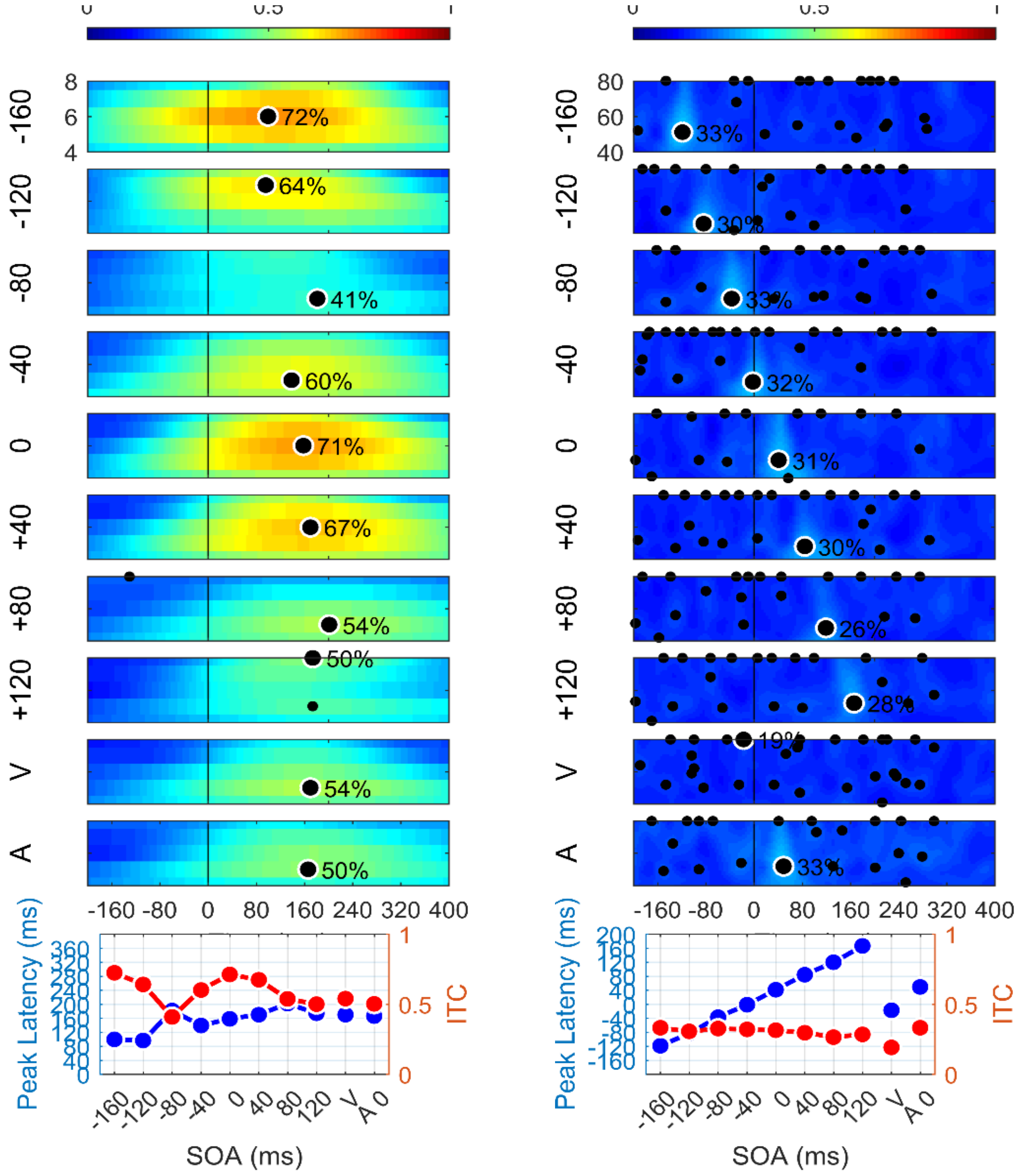


Fig. 25. Time-frequency graphs of intertrial coherence values in low (4~8 Hz, left panels) and high (40~80 Hz, right panels) frequency ranges in posterior channels. Black dots illustrate the peak values, while the one with white circle is the one with the greatest coherence level. Plots given below the graphs illustrate the latencies (blue) and ITC levels (red) of the peaks at the given SOA.

Finally, together with phase analyses, peak power and latency estimations show that unimodal stimulus evokes high-frequency potentials. However, low level evoked potentials have very high promising information about the multisensory integration in posterior channels. Frontal channels have important peak latency changes in a more like linear fashion which correlates positively with RTs.

3.4 Discussions

This experiment was an EEG investigation of audiovisual integrations with the behavioral measurement of the reaction times to visual stimulus. Auditory stimulus was presented in such a wide range of SOA values for the first time in an EEG study of this task. Both additivity analyses and power and phase component analyses of time-frequency decompositions made important contributions to the literature. Although a linear change is observed in reaction time as a function of SOA, additivity analyses revealed that integration mechanisms were different according to the leading modality. This experiment revealed that posterior and the frontal regions play qualitatively different roles according to the sense of leading the modalities. Since quantitative changes in temporal offset lead a qualitative change at 0 ms SOA condition. Furthermore, amounts of additivities in the parietal channels predicts the amount of reaction time shifts for auditory leading conditions. Time-frequency decompositions of posterior channels revealed that theta frequency oscillations were highly correlated with the levels of audiovisual integrations.

3.4.1 The oscillatory pattern given above change in the reaction time to early auditory to late bimodal stimulus

Previous research shows that auditory stimulus speeds up the responses to visual stimuli. Oscillatory changes in the reaction times as a function of SOA for individual participants can be observed, which are in line with the previous studies suggesting the idea that phases of the ongoing responses affect the reaction time to the forthcoming signals (Diederich *et al.*, 2012).

3.4.2 Separate networks for temporal audio interference depending on the leading sense

It was reported that there was an asymmetry in processing mechanisms of audiovisual integration, according to the leading modality (Cecere *et al.*, 2017). Another study also reported that a sensitivity improvement effect was observed in AV simultaneity judgment task when participants were trained with visual leading auditory (VA) stimulus and with both VA and AV stimuli, however, AV training did not lead a significant improvement in judgment (Cecere *et al.*, 2016). In their behavioral study, Cecere et al. (2016) pointed out the different mechanism of the integration, and their EEG study also showed that neural mechanisms were different at specific time windows (Cecere *et al.*, 2017). Although the time intervals where the differences were salient and identified, due to their analysis method, spatiotemporal neural dynamics were not investigated in that study. Therefore, the reasons for the asymmetry in training effect remain unexplained. In this experiment, results in additivity analyses point that auditory leading audiovisual responses interact specifically in the parietal region at the very early time windows of the visual processing. However, when the auditory stimulus succeeds the visual stimulus, the interaction was not observed in the parietal region, which suggests no interaction in the early visual processing areas. On the other hand, in the sound following flash conditions, interactions occur prominently at frontal channels while no interaction observed at the early stages of the visual processing.

In brief, when auditory neural responses arrive at cortex it prepares the occipito-parietal regions quickly for the forthcoming visual signals. When the visual signal arrives at the early visual cortex, visual responses interact with the auditory ‘effects’ at these early regions, but no further interaction was observed at frontal regions. This interaction mechanism might be such robust that training does not change the perceptual outcomes. On the other hand, when auditory responses arrive later than visual, it cannot ‘catch’ the visual responses at occipital regions, however late auditory responses affect the way how visual signals are processed in the frontal regions. Furthermore, possible neural mechanisms of this frontal interaction are such malleable that training may change the perceptual outcomes. Further neuroimaging studies may reveal the neural mechanisms of this malleability in the frontal region. Therefore, these findings may explain the asymmetry in the training effect.

These results are supporting the idea that the auditory neural responses carry some informative signals about the events in the environment that prepares the occipito-parietal region to get it ready for coming visual stimulus and have the visuomotor system to react to it accordingly (Petro *et al.*, 2017). This mechanism is very important for the cases when visual attention is not directed to the event. In these cases, events can take the visual attention via auditory cues and visual system can recognize the event or object. On the other way around, sound succeeding light cannot have a role in the early steps of the visual processing, rather it may strengthen the visual processing in the later steps. Associations of the auditions and vision are also very important for being alert for the forthcoming anticipated events (Stekelenburg & Vroomen, 2012; Thorne & Debener, 2014).

3.4.3 Role of the amplitude and the latencies of the low-frequency oscillations in audiovisual integrations in posterior regions

Time-frequency analyses identified three distinct frequency bands. Auditory stimulus elicits gamma frequency activity and visual only stimulus-evoked low beta activities over the posterior region. Furthermore, they also evoked theta frequency activities. Multi-sensory stimuli evoked these three frequency components; however, theta frequency component was significantly modulated as a function of SOA values. Previous studies indicated that the phase synchrony had an important role in communication between primary sensory areas (Naue *et al.*, 2011; Thorne *et al.*, 2011; Thorne & Debener, 2014). For some regions such as left anterior or right posterior regions, higher beta frequency means faster reactions for AV-0 and for unisensory stimuli (Senkowski *et al.*, 2005). However, SOA ranges were so small in their studies that it was hard to see a general picture of the oscillatory changes.

In general, our analyses suggest that we are coded in early-stage visual sensory centers at amplitude timing in the theta frequency band and support the hypothesis of the research proposal (Hypothesis 3: "Adjusting the timing/duration of the neural response to auditory timing"). We think that early-time window interactions in the occipito-parietal region originate from feedforward crosstalk in the visual and auditory areas, feedback from the central and frontal multisensory fields in the late interaction.

3.4.4 Evaluation of audiovisual simultaneity as a multistage process

If the auditory stimulus is too early, the sound works as a sort of alerting mechanism, and if the sound comes close to the flash in time, it is likely that early region integration is taking place because of the fast connections. Therefore, for the first case, it seems reasonable to hypothesize that feedforward integration takes place because feedforward is too late for an interaction., whereas in the second case, feedback mechanisms are active.

3.5 Conclusion

In this EEG experiment, using an extensive analysis, we investigated the neural mechanisms of temporal integrations of audiovisual stimuli. Our findings were in two-fold. First, occipital and frontal time-frequency decompositions revealed that low-frequency power amplitude and latencies changed as a function of SOA values in audiovisual integration. Second, the timing of the auditory stimulus affected the spatiotemporal profile of the neural dynamics of the integration. When the sound was presented earlier, interactions were located in the parietal region as early as the feedforward signals were processed and timing of the earlier ERP peak predicted the reaction time. However, when sound came in synchrony or later, the interaction was located in frontal regions. In addition, although stimuli were presented in a range of various SOAs, interactions occur in three distinct time windows both in frontal and occipital regions. Recruitment of the early parietal regions in case of sound-leading conditions may represent a mechanism that serves as an alarm for visual processing structures to be prepared for a new stimulus coming. In general, this happens when we first do not look at a specific object: we first hear the sound, and sound takes our attention on it. On the other hand, the sound coming in sync or later may have a different role in higher mechanisms such as helping in the interpretation of visual events when we already look at them. Although our design did not allow to detect them, in both cases preceding modalities may carry some specific information that has a role in having more accurate perception.

CHAPTER 4

4 EXPERIMENT 2. AUDITORY INTERACTIONS IN THE VISUAL TIME INTERVAL PERCEPTION

4.1 Introduction

The second experiment was designed to investigate the neural mechanisms of visual time interval perception with the temporal modulation of accompanying auditory stimulus. Temporal ventriloquism studies suggest that auditory click can change the timing of visual stimuli (Fendrich & Corballis, 2001; Morein-Zamir *et al.*, 2003; Recanzone, 2003). Following this phenomenon, when flashes were presented twice in time, this repetition defines a time interval. If the clicks were presented just before the first flash and after the second flash, then these auditory clicks create a time interval between flashes (Getzmann, 2007; Freeman & Driver, 2008; Kafaligonul & Stoner, 2010, 2012). Therefore, with the fact that sounds can attract light in the temporal dimension, the auditory time interval can change the perception of the visual time interval.

This experiment aims to understand the neural mechanisms underlying temporal ventriloquism and its influences on time interval perception. Specifically, the mechanisms of the multisensory interactions are suggested to be carried out at the early levels of the unisensory processing, and then they are integrated neural signals are traveled to the later information processing areas of the brain. Since time interval between one stimulus and the following must be counted with a mechanism that holds the time passed. Furthermore, to have a measure of the time interval, a reference interval should be introduced and this requires a memory process. Eventually, there should hold two kinds of time information for time interval judgment mechanisms. It is apparent that time interval perception requires more computationally complex processes, compared to the perception of a single event.

4.2 Design and apparatus

4.2.1 Participants

Fifteen healthy adults (5 females; mean age of 22.7±3.7 SD years; 19-30 years range) participated in this 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 Ankara University.

4.2.2 Apparatus

The software, monitor, insert earphones, calibration devices, hardware, and conditions in the experiment room were the same as those of in the first experiment.

4.2.3 Stimuli and Procedure

Spatial and temporal configurations of the flashed bars and auditory clicks were the same as those in the first experiment. Brightness values of the screen and the stimuli were also the same as in the first experiment.

The basic idea in this experiment was to pin down the role of the secondary modality (auditory) in a percept (of time interval) of another primary modality (visual) by systematically varying a property (time interval) of the secondary modality as a factor, while keeping stimulus in the primary modality the same. As this experiment focuses on the perceived time interval of a visual event, two consequent flashed bars was shown as the primary modality. The physical time interval of the auditory clicks was modulated, as shown in Fig. 26, for the secondary modality.

For the purpose given above, two visual flashes were shown for 50-ms during the experiment and a 100-ms time interval (ISI) was used between these two visual flashes. The time interval was defined by visual flashes shown at the same location twice either on the right-side or left. To engage the participants in the stimulus, visual time intervals were presented twice sequentially, and the participants were asked to judge the time interval between these visual events. Participants decided whether the interval appeared on the left or on the right was shorter, and responded by keyboard press. The auditory time interval was introduced with the clicks accompanying the flashed bars. Time intervals of the click were either shorter or longer than visual intervals. However auditory intervals were two level: mild-shorter (ISI: 100-ms), -longer (ISI: 160-ms) and extreme-shorter (ISI: 40-ms) and -longer (ISI: 220-ms) intervals were introduced (Fig. 26). Therefore, there were two AV conditions where participants compared either mild-shorter and mild-longer or extreme-shorter and extreme-longer auditory modulations, while the visual intervals were the same (ISI: 100-ms) in both the first and second intervals. As in the first experiment, to use in additivity analyses, one visual-only and four auditory-only conditions (extreme- and mild-, -shorter and -longer) were also introduced. Therefore, we presented both visual and auditory stimuli in a modality-specific way and in combination. Participants were asked to respond only to the visual stimulus and not to do anything in auditory only conditions. Either short or long auditory time interval was presented pseudo-randomly one after another at a temporal offset (ISI) of 900-ms.

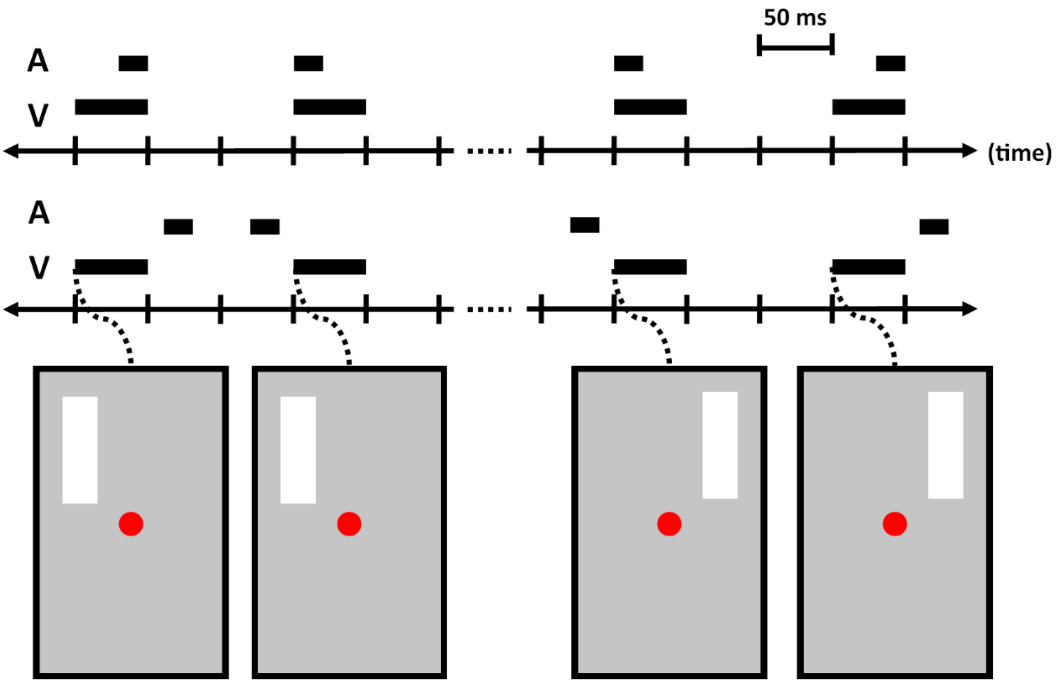


Fig. 26. Audiovisual conditions that are used in visual time interval experiment. Demonstrations of the trials for mild (top) and extreme (bottom) auditory modulation conditions at the temporal dimension. Short (left) and long (right) auditory modulations are given in random order. The bar is presented twice first either on the left or right randomly and in the second time presented on the contrary side. Boxes at bottom illustrate how the visual stimuli change in time.

Therefore, 4 audiovisuals, 4 audio-only, 1 visual-only and 1 catch trials have been presented within 4 blocks. In each block, trials were presented 12 times, randomized in terms of location (first double flash on left or right) and temporal order of shorter and longer auditory conditions (first auditory interval was short or long). Eventually, every condition had 48 trials. Participants were encouraged to have a short break (generally less than one minute) between blocks to maintain high concentration and prevent fatigue.

Moreover, to ensure that participants perceive the visual time interval difference of the stimulus a practice session has been prepared. Participants who did not discriminate the visual time intervals clearly in the same visual setup did not participate in EEG experiment. Participants compared the 100-ms ISI interval with 40- to 190-ms visual time intervals, with 30-ms steps, in this practice session. In addition, to catch if participant really attends to the visual stimulus, we included catch trials in the experiment, where the auditory ISI was the same in reference and test stimulus, however, the visual stimulus differed with 30-ms SOAs, which means that there were 120-ms interval difference in two consecutive visual time intervals.

Only the participants who had more than 70% accuracy at catch trials were included in the analyses.

In the time course of a trial, after 400-ms pre-stimulus interval, a visual bar flashed twice either on the left or right of a central fixation point pseudo-randomly. As shown in Fig. 26, the double-flash sequence was presented twice sequentially, however, the second flash is presented on the contrary side after 900-ms (a reference to test sequence interval). The reaction was expected within 800-ms, otherwise, the trial was ended and continued to the next trial. The inter-trial interval was randomly varied from 500 to 1500 ms.

4.2.4 EEG preprocessing and analysis

EEG preprocessing steps were the same as in the first experiment. Like in the first experiment, event marker was set on the first visual stimuli onset of the time intervals. Eventually, there were 4 **AV**, 4 **A** (extreme-shorter and -longer, mild-shorter and -longer), and 1 **V** conditions. Catch trials were excluded from EEG analyses. Therefore, the locations (left, right) and the order (reference, test) variables were combined.

After applying standard preprocessing procedure, average ERPs have been prepared for each participant for all conditions, except catch trials. As there were 2 independent variables, the degree of modulation: mild and strong; and length of modulation: shorter and longer, for main conditions have been considered. In the time-course of one trial, two time intervals have been presented sequentially. Therefore, in ERP analyses, all trials of a condition have been averaged, regardless of the temporal order in the presentation order or of the location relative to the fixation point.

Like the first experiment, in ERP analyses additivity hypothesis have been tested. For this purpose, both unimodal auditory, visual, and multimodal audiovisual evoked potentials have been recorded in with the same temporal structure.

$$\boldsymbol{Additivity = AV_{SOA_i} - (A + V)_{SOA_i},}$$

$$\boldsymbol{where\ (A + V)_{SOA_i} = \left(A_{SOA_i}\right) + V, and\ SOA_i = -45, -15, 15, and\ 45\ ms}.$$

The essential approach in ERP analyses was to find sub- or super-additive influence of audio evoked neural responses to visual responses. $\boldsymbol{A}$ and $\boldsymbol{V}$ responses were linearly added and summation ($\boldsymbol{A + V}$) was compared with $\boldsymbol{AV}$ responses and additive effects found in all channels. According to the additivity in channels and time, representative channels have been selected to be averaged from specific regions of interest or further analysis. These channels were O1, Oz, PO3 from occipital; F3, AF3, F1 in frontal areas.

For the analysis of evoked potentials, after preprocessing and averaging the trials, 40 Hz low-pass filter is applied for smoothing the ERPs. Before comparing the $\boldsymbol{AV}$ ERPs with $\boldsymbol{A}+\boldsymbol{V}$ ERPs $\boldsymbol{AV}$ ERPs were baseline corrected to the time window between -150 to -50 ms before the first visual onset. ERPs are compared with paired running *t*-test. In order to clear noisy significant points, the minimum sequential 10 significant points were kept as significant as described in the experiment 1.

4.2.5 Time-frequency analyses

To get possible event-related oscillatory patterns, wavelet transform has been applied over the ERPs. The method for the time-frequency decompositions was the same as in the first experiment. The average evoked spectral amplitudes have been calculated by first averaging all trials and then transformed into the time-frequency domain. Different frequency bands have been defined (also in accordance with the first experiment). Then, within these ROIs, average baseline evoked activity was calculated from -200 ms to -100s for each participant and the time duration for evoked activity was defined. Durations were started from the beginning on at 10 consequent points which were 3 × SD bigger than the baseline value and ended with the 10 consequent points smaller than baseline plus 3 × SD. The amplitude of a specific time and frequency is assumed to be a direct indicator of the evoked activity that includes only phase-locked oscillations (Fig. 27). Mainly two spectral features were extracted from time-frequency decompositions. First, *rising* from and *falling* down to *baseline* points have been calculated, since, these points have been used for calculating power increase *duration*. However, this measure did not lead a stable result. Alternatively, peaks have been estimated and *half peak* values were also used in order to calculate edges for both rising and falling sides of the power hill. Likewise, average amplitude within the 'amplitude hill' window was calculated as a second metric.

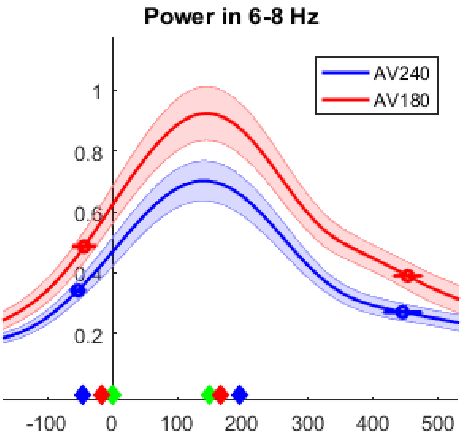


Fig. 27. Sample window edges which were depicted by the points where the power rises from baseline and fall back to the baseline (see big dots on the lines). The time interval between these two points was defined as the duration of the power increase. Likewise, average power within this interval was defined as the average power of the evoked oscillation.

4.2.6 Behavioral data analysis

For behavioral data, we only used trials included in the EEG analyses. We calculated average performance across participants for each stimulus condition. To determine whether the effects of SOA significant, we applied pairwise comparisons this behavioral measure. A simple linear regression was calculated also to predict RT based on SOAs.

4.3 Results

In the present experiment, we used high-density EEG to investigate the neural correlates underlying auditory influences on visual time interval perception. We recorded ERPs during a classic two forced-choice tasks in two **AV** conditions and one **V** condition and used additivity analysis on ERPs and analyses on spectral amplitudes component of time-frequency decompositions. Under the hypothesis that different neural networks and mechanisms are involved in audiovisual temporal binding depending on SOAs, quantitatively different spatiotemporal interactions patterns are expected to emerge across different SOAs, specifically, amplitude and durations were investigated systematically.

4.3.1 Amount of auditory modulation influences the visual time interval perception

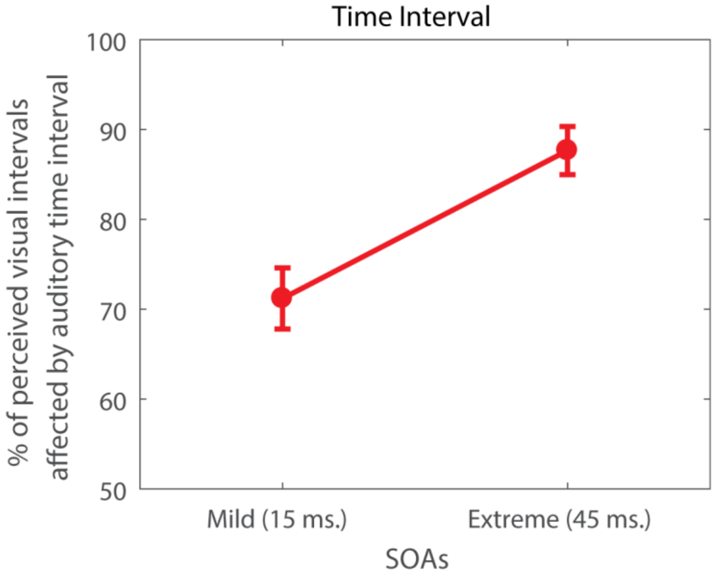


Fig. 28. Display of percentages of the visual time interval perception affected by the simultaneous auditory time intervals. Extreme auditory modulation affected time interval perception of visual stimuli more than mild modulation, also the effects were above the chance level of %50.

In visual-only condition, in which two exactly the same visual time intervals were compared, and participants perceived first and second time intervals equally long, at chance level 50% (M = 53.33%, SD: 21.14%); $t_{14} = 0.61$, $P = 0.55$. As shown in Fig. 28, the visual time intervals modulated by shorter sound intervals was perceived significantly shorter compared to the ones that were modulated by longer sound intervals in both mild modulation condition (M = 71.2, SD: 13.16); $t_{14} = 6.24$, $P < 0.001$, but especially in extreme modulation comparison (M = 87.7, SD:10.37) condition; $t_{14} = 14.07$, $P < 0.001$. When SOA was extreme (45 ms), the percentage of perceived visual interval affected by auditory interval was significantly higher compared to the mild modulation case (SOA: 15 ms); $t_{14} = 5.78$, $P < 0.001$.

4.3.2 Auditory time interval effects on ERPs elicited by visual time interval: Various regions and time intervals

We wanted to determine how distinct auditory time intervals presented simultaneously to visual time interval affect evoked activities to a visual time interval (i.e. the visual test stimulus which is the same across all conditions). As behavioral results showed, various auditory time intervals affect the visual time interval perception. Therefore, we reasoned that we should observe significant differences in the visual components of the audiovisual evoked potentials. Moreover, if interactions occur at the early steps of the visual processing, they should be observed over early visual cortex. Otherwise, interactions could also be observed at locations where the time interval is processed. To test these possibilities, pointwise running repeated measures ANOVA (with basic corrective steps) was performed for different auditory time interval conditions and mild and extreme modulations as within-subject factors. Furthermore, for each level of the factors, levels of the other factors were also compared with a pairwise running *t*-test. In other words, comparisons of extreme and mild for shorter and longer conditions and shorter and longer for extreme and mild conditions were held. These additional comparisons allowed us to reveal the spatiotemporal profiles of the ERP interactions. As shown in Fig. 29A, there was no salient interaction effects. However, short and long auditory time intervals caused differences mostly observed over frontal regions. Further *t*-test comparisons of extreme short and long auditory time intervals; and mild short and long auditory time intervals had different spatiotemporal profiles (Fig. 29B). Extreme short and extreme long auditory stimuli comparison showed an effect on visual processing at the time window of 50-80 ms, mostly at the right frontal region. However, mild short and long effect on visual processing was observed mostly at 200-220 ms time window, at the frontal region; and very early left centroparietal differences were also observed. On the other hand, comparisons of mild and extreme auditory modulations on visual processing were different in longer and shorter. Although there was no difference between mild and extreme auditory modulations on visual processing for longer auditory stimuli; for shorter stimuli, a difference was observed at 190-220 ms, distributed almost everywhere in scalp topography. Importantly, on the other hand, there were very little channels and time intervals where significance observed after 250 ms, which suggests that the auditory signals influence the visual time interval processing already at early sensory processing steps, in the context of the experimental factors.

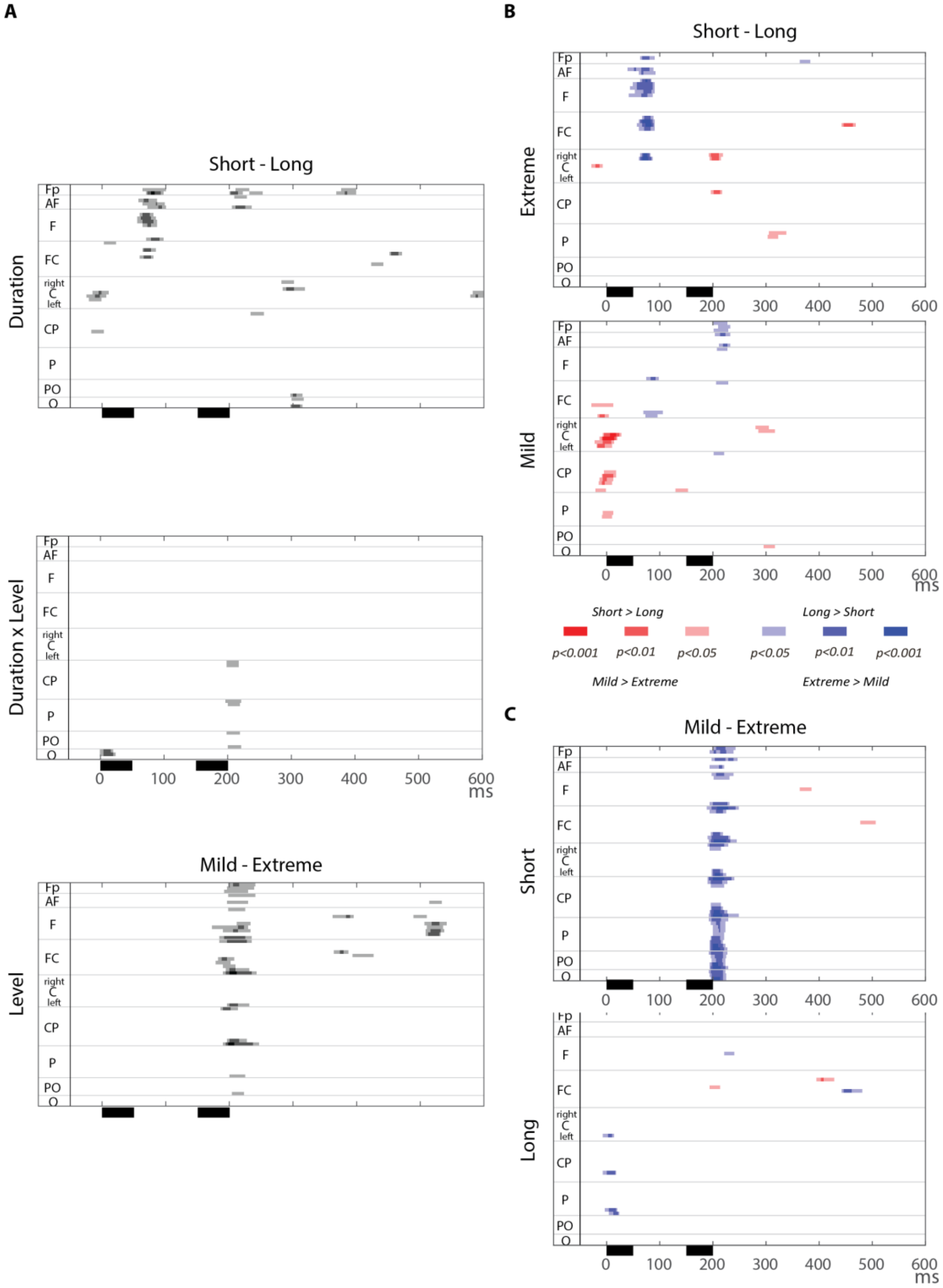


Fig. 29. Time courses of auditory effects on visual time interval evoked activity. (A) Running repeated-measures ANOVAs with level (mild and extreme), an auditory time interval (shorter and longer) and the interaction of these variables as factors. A specific time point was shaded only if at least 20 ms of contiguous data had significant two-way interaction (modality × time interval). (B) Pointwise running *t*-tests on the evoked activity to time interval modulated with shorter and longer

auditory stimuli. The pairwise (shorter and longer) comparisons for each level are shown in separate plots (mild: upper panel, extreme: lower panel). In each plot, time is displayed on the x-axis from -50 to 600 ms (relative to the first flash onset), and electrodes are displayed on the y-axis. The significant difference between time interval conditions is marked on the temporal axis by different levels. A specific time point was shaded only if at least 20 ms of contiguous data was significantly different (see Methods). Based on the sign of the difference between shorter and longer time interval conditions, either red (shorter > longer) or blue (shorter < longer) color was used for shading. Flash frames defining time interval are represented by black rectangles at the bottom of the timeline. The timing and duration of flash frames are indicated by the position and length of these rectangles, respectively. Other conventions are the same as those in the Fig. 17.

EEG analyses have been carried out to find out how the perceptual effects -revealed by behavioral analyses- overlap with the oscillatory responses in the different regions of the cortex. Fig. 30 shows the average ERPs for **AV**, **A** and **V** conditions from exemplar occipital, and frontal ROIs (for the additivity analyses at all channels of all conditions, see Fig. A1). The visual-only stimulus was elicited a significant activation in the area above the early visual cortex, and these locations were activated also for various auditory-only conditions, but with a smaller amplitude. As for the frontal channels, auditory responses were more salient in this region compared to the posterior regions. In the posterior region, V stimulus elicited, as expected, P1, N1, and P2 peaks clearly at around 120 ms, 180 ms, and 290 ms, respectively; and their activities started at 50 ms, 140 ms, and 200 ms, respectively. Time windows in which the interactions occur nicely fit in these time windows framed from the start of the activities to the point where the activity reaches its peak. Most of the time windows in which differences were significant are located at those ranges starting with the activity onset and to the peak of the activity; in other words, interactions were observed at the rising edges of the peaks. First, from the results of the **AV** and **A+V** comparisons only for the mild shorter condition, a significant super-additive effect was observed at 200-220 ms (Fig. 30A). Similarly, also for the frontal region, it was hard to describe an interaction pattern; however, a super-additive interaction was observed at 220-250 ms time window (Fig. 30A). In addition, due to the reason that frontal region was closer to the auditory cortex, auditory activities were more salient in this region compared to posterior regions. In general, although some minor interactions were observed in several locations and time points, as Fig. A1 shows, it was hard to identify a pattern distributed over conditions. Interestingly, in almost all conditions there is no any identifiable additivity over the occipital cortex. As can be observed in Fig. 30A, ERP components from the posterior region are reasonably like the ones in Experiment 1, namely where the first positive peak responded to visual stimulus is observed at 100 ms. ERP waveform goes as it does in the first experiment and at around 250 ms, differences can be observed due to the second flash, and modulation of the auditory stimulus can also be observed in this second peak in amplitude. However, there is no significant additivity at any time point at this posterior region, even though significant additivity effects were observed in the single click and single flash experiment. Similarly, also in the frontal channels, it is hard to identify a pattern distributed in all conditions which do not lead a story about interaction.

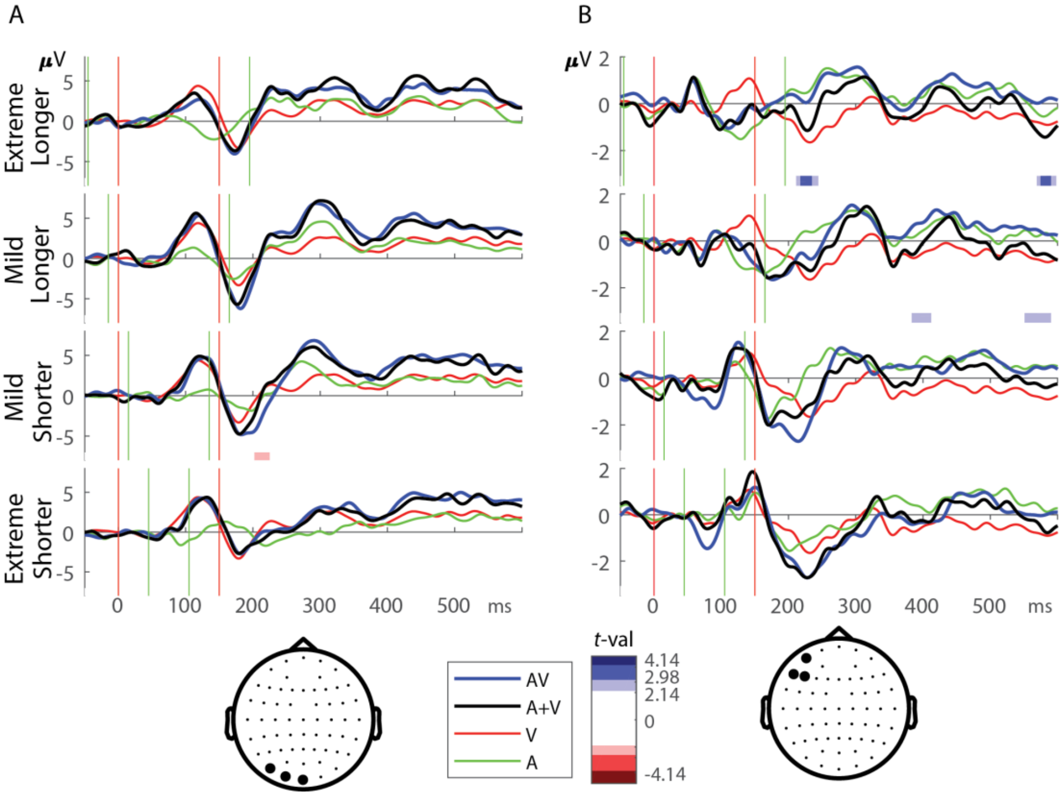


Fig. 30. ERP plots for time interval judgment experiment and **AV** and **A+V** comparisons for (A) occipital and (B) frontal channels for all click conditions. Both extreme (SOA= ±45) and mild (SOA= ±15) auditory time interval conditions can be observed separately. Red and blue bars under the plots indicate the level of significant differences between **AV** and **A+V**. As the results of *t*-tests for the corresponding points on the timeline indicates, red colors represent significant sub-additivity, while blue is for super-additivity.

As the participants compared the visual time intervals which are manipulated by shorter and longer auditory intervals, we compared both (**AV-A**)s and **AV**s for shorts and longs (see Fig. 31). In the occipital area, there was no significant difference, but in frontal area, at around 210 to 235, short click had a bigger peak amplitude than long click. As for **AV** conditions, as expected, due to the auditory component persistence, there were more significant points, and extremely short and long click causes more significant points, especially after 200 ms.

In conclusion, although additivity analyses did not indicate a pattern distributed over different conditions, **AV-A** comparisons of the experimental conditions revealed interesting spatiotemporal profiles which are in line with the previous studies (Kaya *et al.*, 2017).

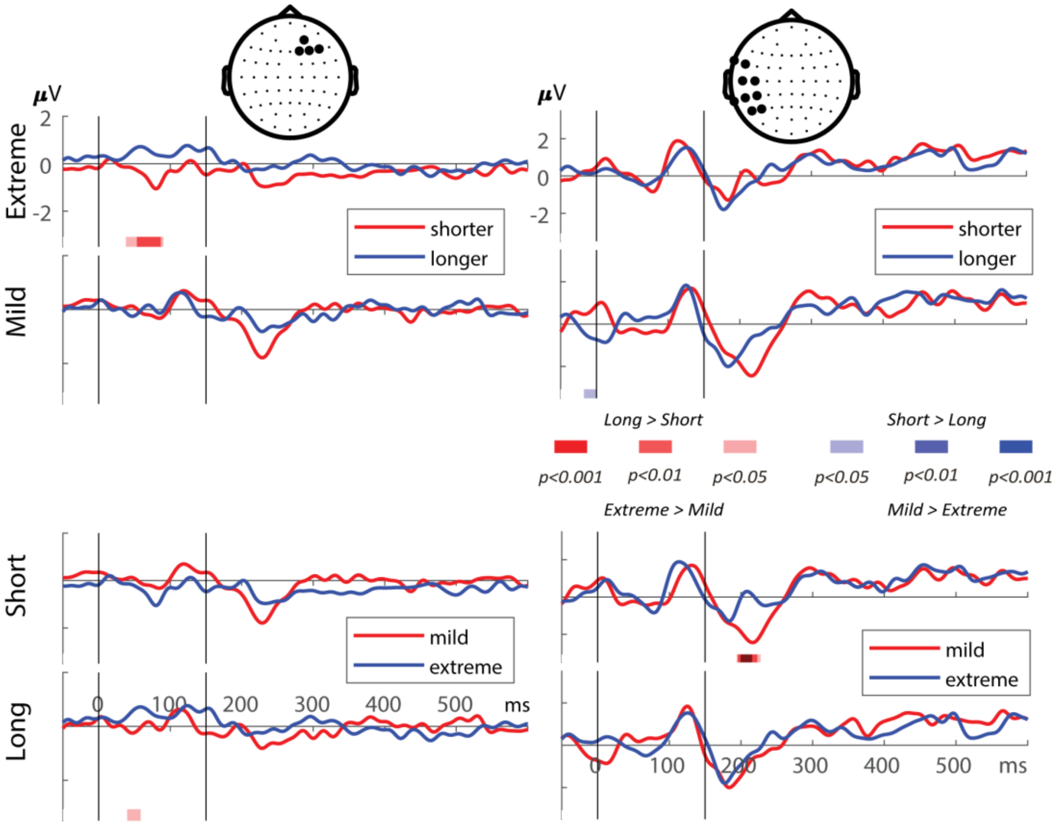


Fig. 31. **AV-A** ERPs comparisons for shorter and longer auditory modulations of visual time intervals, separately in mild (panels in the first and second rows) and extreme (in the third and fourth rows) levels at (A) right frontal and (B) left temporal regions. The time point where the difference is significant were marked by colored bar below the ERPs. Tones of the color indicate the *p* values of the *t*-tests which correspond the levels of *P* = 0.05, 0.01 and 0.001.

4.3.3 Oscillatory evoked-power of the frontal ROI changes as a function of the direction of the auditory time interval while occipital is to the amount of the SOA values.

As visual and auditory stimuli create responses, we expected to see an increase in amplitude and duration of the oscillatory power which was related to audiovisual interaction. To understand the oscillatory dynamics of this experiment in detail, wavelet transform of the signal have been obtained and afterward, some features are extracted from the time-frequency transforms. Spectral maps from two exemplary ROIs from frontal and occipital regions revealed that two frequency bands had high power components, 4-8 Hz and 8-12 Hz. Power and duration analyses were applied to these frequency bands. As revealed in

Fig. 32 (panels in the first column, second and third rows), Fig. 33A and B (panels in the first column and second rows), mild interactions of shorter and longer auditory stimuli created more power for AV conditions in the occipital regions at 4-8 Hz band ($F_{1,13}$ = 35.626, P = 0.000, η_p^2 = 0.733). In addition, a two-way interaction was observed in a way that shorter auditory stimulus elicited higher compared to longer power when modulation was mild, on the other hand, longer auditory elicited higher power when modulation was extreme ($F_{1,13}$ = 11.740, P = 0.005, η_p^2 = 0.475). However auditory-only spectral maps reveal that this difference was related to the auditory component which has the same effect of SOA level ($F_{1,9}$ = 5.277, P = 0.047, η_p^2 = 0.370). When the auditory component was subtracted from the AV responses, AV-V responses were similar to the visual-only responses.

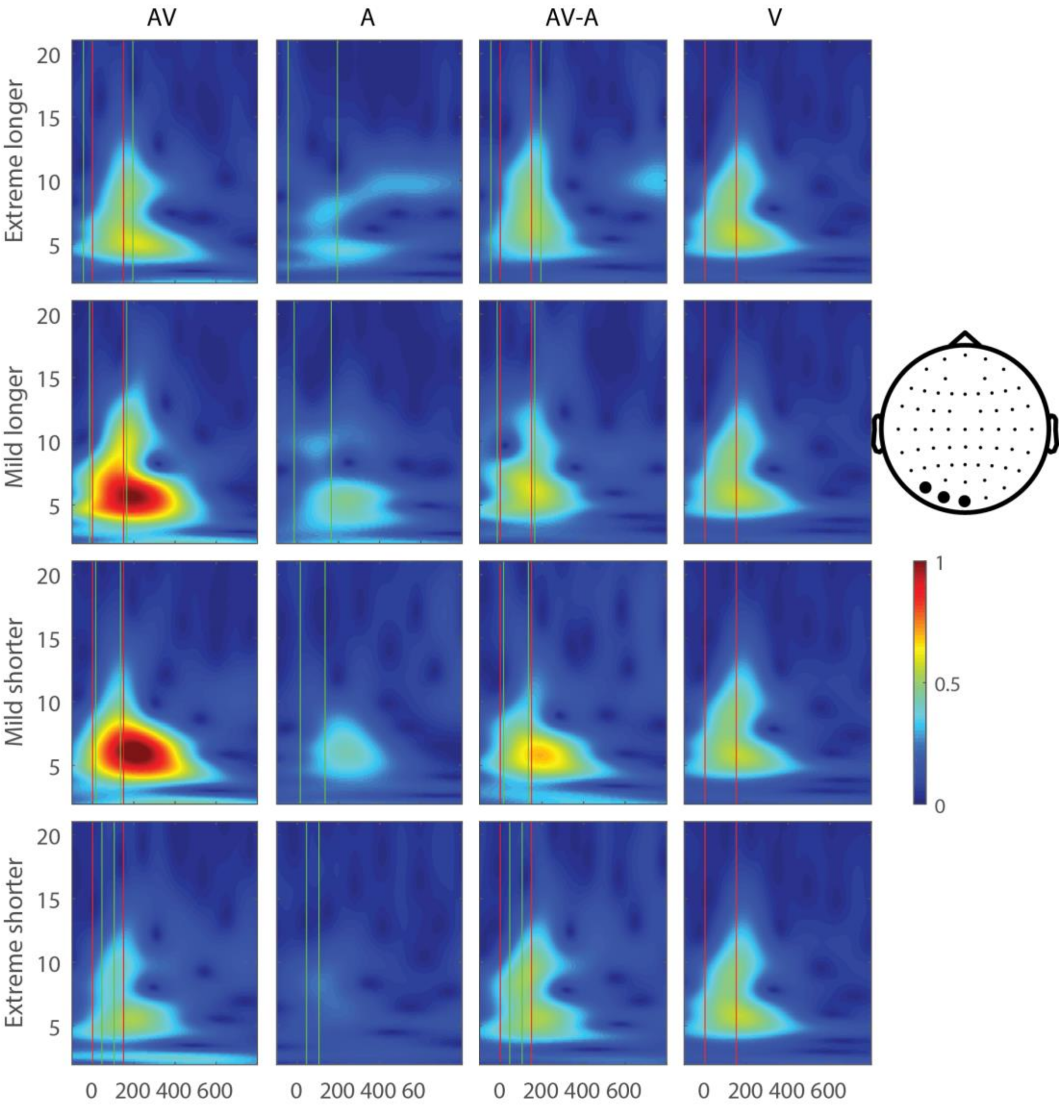


Fig. 32. Three occipital channels given in the topography were averaged for calculating the time-frequency decompositions on average AV, A, V, and AV-A modalities. The vertical red lines depict visual stimulus time, green lines are for auditory. Spectral maps of the responses for various modalities were distributed in columns while experimental conditions were presented in rows.

Duration was also differed between mild and extreme SOA values in the way that extreme modulation caused longer duration responses ($F_{1,13}$ = 8.431, P = 0.012, η_p^2 = 0.393); however, longer and shorter stimulus did not cause a difference at occipital region. To sum up, 4-8 Hz frequency maps from occipital region is sensitive to the level of modulation in both power and duration features of the responses. The mild modulation caused higher power in occipital ROI in general, and this effect is more salient in shorter-interval auditory modulation compared to longer one. Extreme modulation evoked longer responses at this region on the other hand.

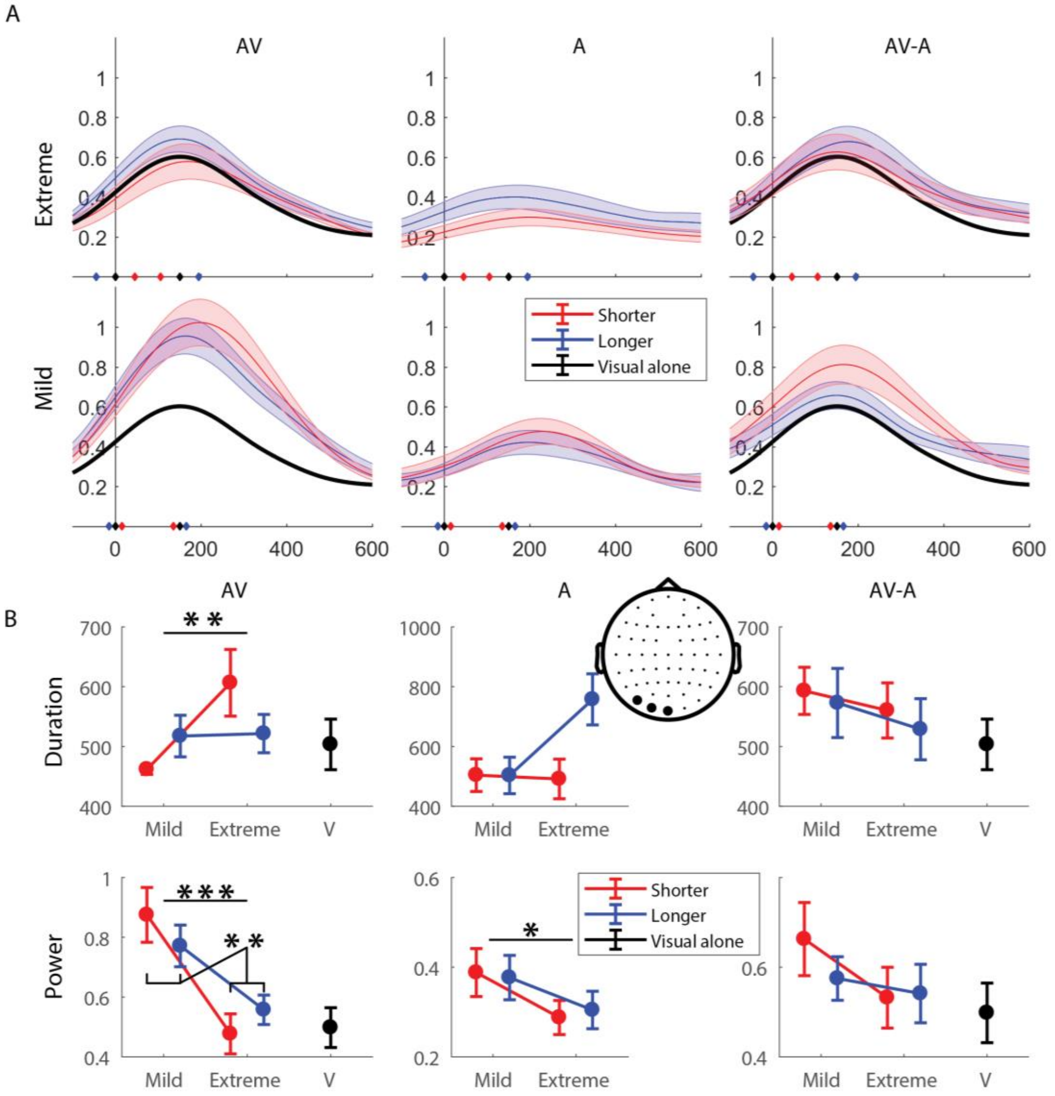


Fig. 33. Oscillatory power comparisons in the left occipital channels ROI for shorter and longer auditory modulations (frequency band: 4-8 Hz). (A) Average powers elicited by the shorter and longer auditory intervals. Shorter and longer conditions were depicted also by the same colors in the plots. Colored dots in the baseline indicate the timings of the visual (black) and auditory shorter (red) and longer (blue) stimuli. Shades around the plot indicate standard error (± SEM) across participants. Mild and extreme modulations were presented in upper and lower panels, respectively. (B) Averaged

oscillatory amplitudes and duration of the amplitude increases for different modalities of different auditory time interval conditions (bimodal: AV, unimodal: A, and synthetic AV-A) (N = 15). In the upper raw, in each plot, the red, blue, and black symbols correspond to shorter, longer auditory intervals, and visual only time interval conditions, respectively. Error bars indicate standard error (± SEM) across participants. A significant difference between shorter and longer conditions, or mild and extreme conditions, or two-way interaction effects of two variables were marked with asterisk signs (repeated measures ANOVA, *$P < 0.05$, **$P < 0.01$, **$P < 0.001$).

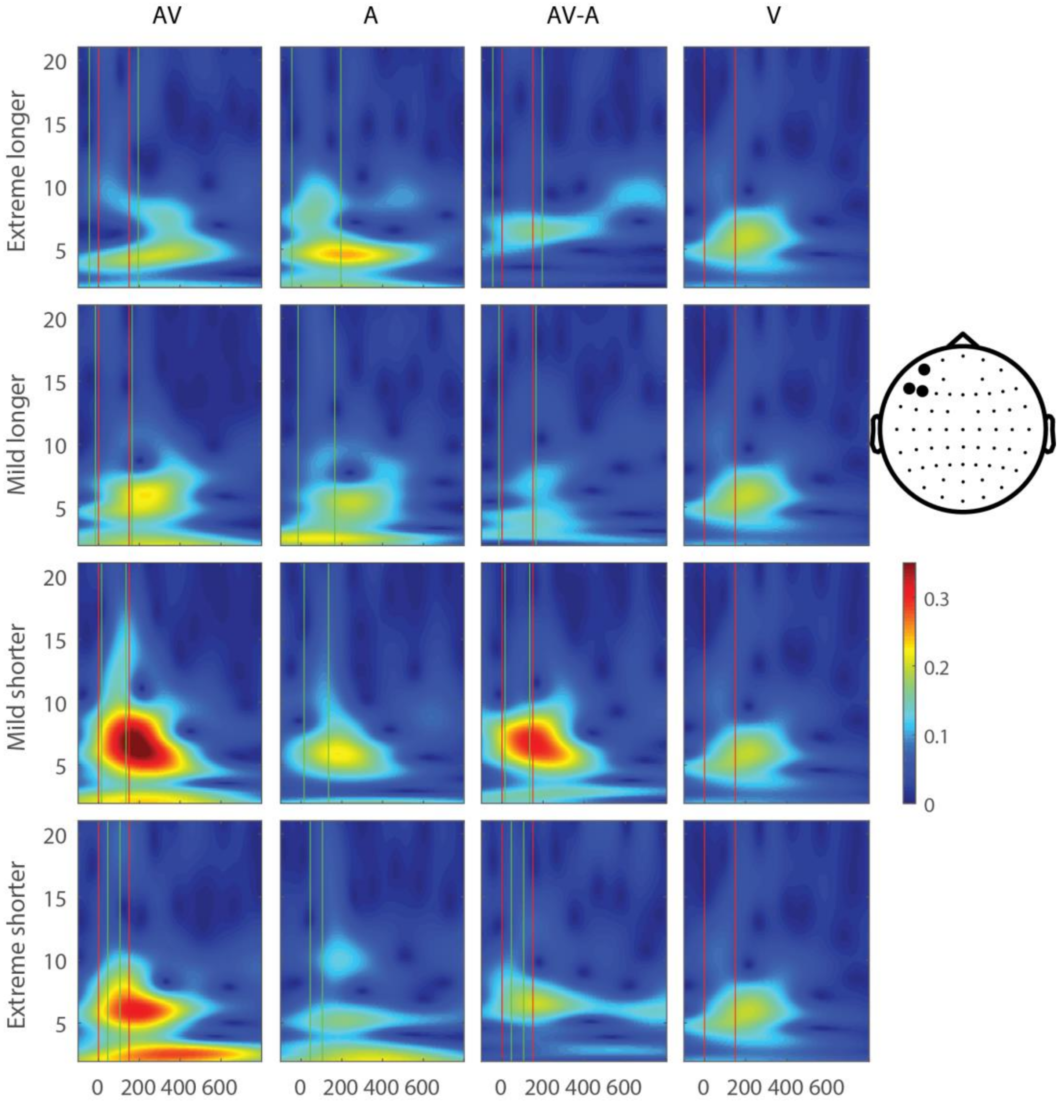


Fig. 34. Time-frequency decompositions on average AV, A, V, and AV-A modalities of evoked potentials to the apparent motion from the left frontal region. Other conventions are the same as those in Fig. 32.

Fig. 34 and Fig. 35 shows that frontal ROIs indicates a different power pattern. Shorter auditory time interval creates more power in both extreme and mild levels of SOA ($F_{1,12}$ = 13.902, P = 0.003, η_p^2 = 0.537). Interestingly, this effect was observed at the AV-A responses, too ($F_{1,9}$ = 8.541, P = 0.017, η_p^2 = 0.487) and there were no significant differences in factors of the responses for auditory only stimuli. there were no effect of auditory duration or modulation level on the duration of the responses. To sum up, power of the high theta neural responses in this frontal region was sensitive to direction of the modulation, even when the auditory component was subtracted.

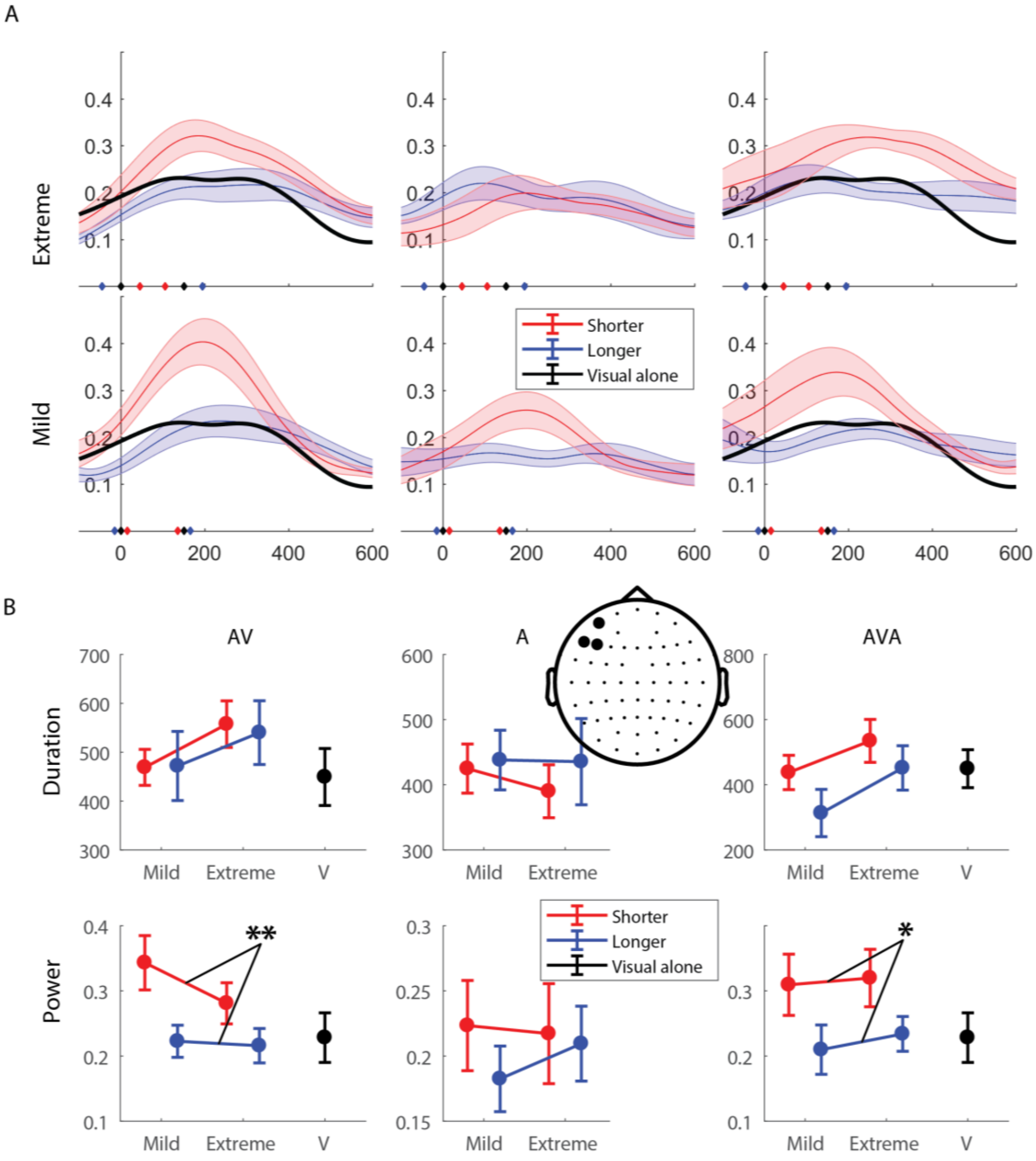


Fig. 35. Oscillatory power comparisons in the left frontal channels ROI for shorter and longer auditory modulations (frequency band: 6-8 Hz). Other conventions are the same as those in Fig. 33.

4.4 Discussions and Conclusions

As the perception of the visual time interval significantly changed with the direction of the auditory stimuli and the percentage of the affected trials significantly changed with the amount of the SOA value, auditory influences on the visual time interval perception was investigated in the neural responses. We hypothesized that auditory influences on the visual processing network could be discovered either at the early step in the visual processing and transferred to the mechanisms where the visual time interval was processed or auditory and visual time interval can be calculated separately and auditory interval dominates the visual interval in time. In both ERP and time-frequency features based analyses on the AV-A responses, duration of the responses induced in the frontal region represented the difference between shorter and longer auditory effects.

Considering possible time interval processing networks, it is highly possible that, we can expect recurrent networks, that is set when the first stimulus arrives and reset when the second event arrives. The question was where the visual time interval is calculated and what is the spatiotemporal profile of the auditory influences on the visual time interval processing. There might be two approaches to solve this question. The first hypothesis was that interactions might occur at the very early perceptual processing networks and each visual event that defines the time interval is transferred -after being modulated earlier- to the mechanisms that calculate the visual time interval. Second, time intervals of auditory and visual stimuli are calculated separately in the modality-specific regions and auditory modality dominates the perception of the visual modality and determines the final perception of the visual time interval.

4.4.1 Irregular patterns of additivity analyses were compatible with the Experiment 1 results

According to the *t*-test results of the additivity analyses within each condition, interestingly, no interaction observed in the occipital area. This might be an indicator of a very small amount of early auditory interaction in the visual cortex in this task. When considering the other regions, and time windows, even if there were interactions at specific locations and time intervals, it was hard to identify a pattern of additivity of any interaction. This might be an indicator of different audiovisual processing mechanisms for each case. No to see a common pattern distributed over all conditions is also in line with the results of the first experiment. As Experiment 1 suggests that the auditory leading and visual leading interactions were different, in all conditions first and the second click-flash couples were different in terms of the modality leading sense. Since, from the perspective acquired in Experiment 1, one flash is modulated in 4 four possible ways and only one of them was auditory leading with 30 ms. SOA, synch, 30 and 60 ms. visual leading. Therefore, not to observe early auditory interaction is expected. This explains the poor interactions in the first flash. In this experiment, the second flash is presented before the evoked activity of the first flash dissolved, therefore some interactions could not be for each click-flash couple.

4.4.2 Comparison of the AV-A responses reveal the differences between short and long; and mild and extreme in various spatiotemporal profiles

As the first experiment confirms, proceeding auditory signals interact with the visual signals at the occipital cortex at early and late time windows. In addition, at the frontal regions, additivity effect has been observed for auditory succeeding conditions, too. Therefore, both frontal and occipital interactions can be an indicator of the early auditory entrainment to the visual processes, if observed at early time windows such as before 200 ms. Results of the ERP analyses were three-fold. First, the comparison of the AV-A responses of the conditions in two dimensions showed early differences due to the auditory influences. Short and long differences were salient in the frontal region: strongly for extreme modulation conditions, as well as for mild modulation. Second, mild and extreme modulation effect was also observed, specifically at the short modulations. This modulation effect was observed in occipito-temporal and frontal regions. Third, there were no late differences in any region. Altogether, these results suggest a strong evidence for the hypothesis that the visual signals were influenced by the early visual processing steps and temporally integrated as early as in these steps.

4.4.3 Occipital channels are sensitive to the level of auditory modulation, while frontal channels are sensitive to short and long modulation

Powers in 4-8 Hz and 6-8 Hz were a good indicator for discriminating the effects caused by audiovisual integration in comparisons of different conditions. Short clicks over the visual time interval, created more power than long ones in the left frontal channels even in AV-A condition. On the other hand, there was no clear pattern for short and long comparison in the occipital channels, however, mild clicks created more power relative to the extreme ones.

Spatial distributions of the comparisons of AV-A responses and time-frequency results in this experiment suggested that the occipital region was sensitive to the extreme and mild modulation. In contrast, the difference between short and long auditory time intervals was more salient in the frontal regions. With the assumption that time interval difference is introduced with the short and long auditory modulations, this difference was processed at the frontal regions, mostly. Similarly, not the direction of the modulation, however, the level of the modulation was introduced by the factor of extreme and mild modulations. Hence, the level of processing was an indicator of audiovisual interaction while the short and long difference was an indicator of time interval perception. As a result, visual time interval judgment is processed in the frontal regions, most probably just after the second visual stimulus is presented, yet the auditory stimulus influences the visual processing directly in each flash independent of the second flash and click.

CHAPTER 5

5 EXPERIMENT 3. AUDITORY INTERACTIONS IN THE VISUAL APPARENT MOTION

5.1 Introduction

Recognition of time perception and the underlying neural mechanisms are one of the most researched areas of neuroscience (Meck & Ivry, 2016). The temporal ventriloquism illusion is one of the most important evidence that audiovisual interactions play a role in shaping time interval perception. In addition, recent adaptation studies have shown that adapting to different dimensions of auditory timing (such as time interval or frequency) significantly affects the subsequent visual timing (Barakat *et al.*, 2015; Levitan *et al.*, 2015; Li *et al.*, 2015; Linares *et al.*, 2016). More importantly, adapting to auditory timing has been shown to affect visual perception (Zhang *et al.*, 2012; Zhang & Chen, 2016). These studies point to several important conclusions. First, parallel to transcranial magnetic stimulation studies (Kanai *et al.*, 2011), their results support the dominant role of the auditory system over other modalities in the temporal dimension. In addition, modalities show the presence of a common temporal mechanism which plays an important role in shaping motion perception.

The third experiment was designed to investigate the neural mechanisms of speed perception, where motion created by visual apparent motion and the visual time interval is modulated by auditory modality. This experiment has very similar design to the second one. We only change the location of the flashed bar randomly from one side to the other in a way that creates apparent motion illusion in one double flash sequence. Like the second experiment, the time interval between two flashed bars is modulated by accompanying stationary auditory ‘click’ sound. Two modulated apparent motion sequences are presented sequentially, and participants had to judge the speed of the sequences, and press the corresponding button which indicates the first or the second sequence.

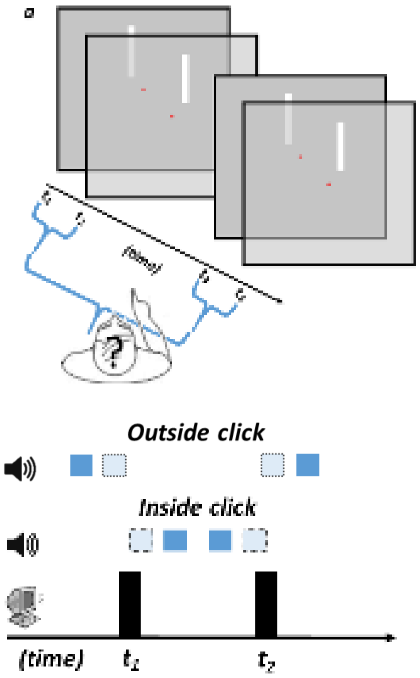


Fig. 36. Experiment design. (A)Two bars were flashed twice in a short time interval with a location change, in a way creating an apparent motion effect. (B) Two clicks defining shorter and longer time intervals accompanied the flashed bars.

5.2 Design and apparatus

5.2.1 Participants

In this experiment, 23 participants participated in the experiment and 4 of them were excluded due to the reason that they cannot focus on the visual event in the catch trials or due to extreme noise in the EEG data. 19 healthy adults (7 females; mean age of 23.5 ± 3.7 *SD* years; age range 19-32 years) participated in the experiment.

5.2.2 Apparatus

The software, monitor, insert earphones, calibration devices, hardware, and conditions in the experiment room were the same as in Experiment 1 & 2.

5.2.3 Stimuli and Procedure

Configurations of the flashed bars and the click sounds were the same as in the second experiment. Brightness values of the visual stimuli were also the same as in the second experiment.

The essential idea in this experiment, similar with the first two studies, was to pin down the role of the auditory modality in the speed percept of visual apparent motion modality (visual) by systematically varying a property (auditory time interval) of the secondary modality as a factor, while keeping stimulus in the primary modality the same. Therefore, this experiment focuses on the perceived speed of an apparent motion, since two consequent flashed bars was shown as the primary modality. As for the secondary modality, physical time onsets of the clicks were same as Exp. 2, as shown in Fig. 36.

In this context, stimulus conventions were all the same as in the second experiment. The only difference was that the location of the second bar was presented on such a contralateral side that it defines an apparent motion. Two exactly same visual apparent motions were presented with the auditory modulations. Auditory and visual presentations times, ISIs, SOAs, time interval between presentations, and unisensory and multisensory conditions were also the same Fig. 37. We asked participants to judge the faster visual motion whether either it was the first or the second, and they indicated their decision by pressing the corresponding button keyboard. Like in the Exp. 2, prior to the actual data collection, to ensure that participants perceive the speed differences between the visual apparent motion of the stimulus a pre-test have been prepared. Participants who did not discriminate the speed differences in the same visual setup did not participate in EEG experiment. Similarly, conditions and the stimulus conventions were kept the same as in the second experiment, however, only the location of the second flash changed in the flash couples that define a time interval, now in this experiment defined apparent motions. The same participation selection criteria were applied.

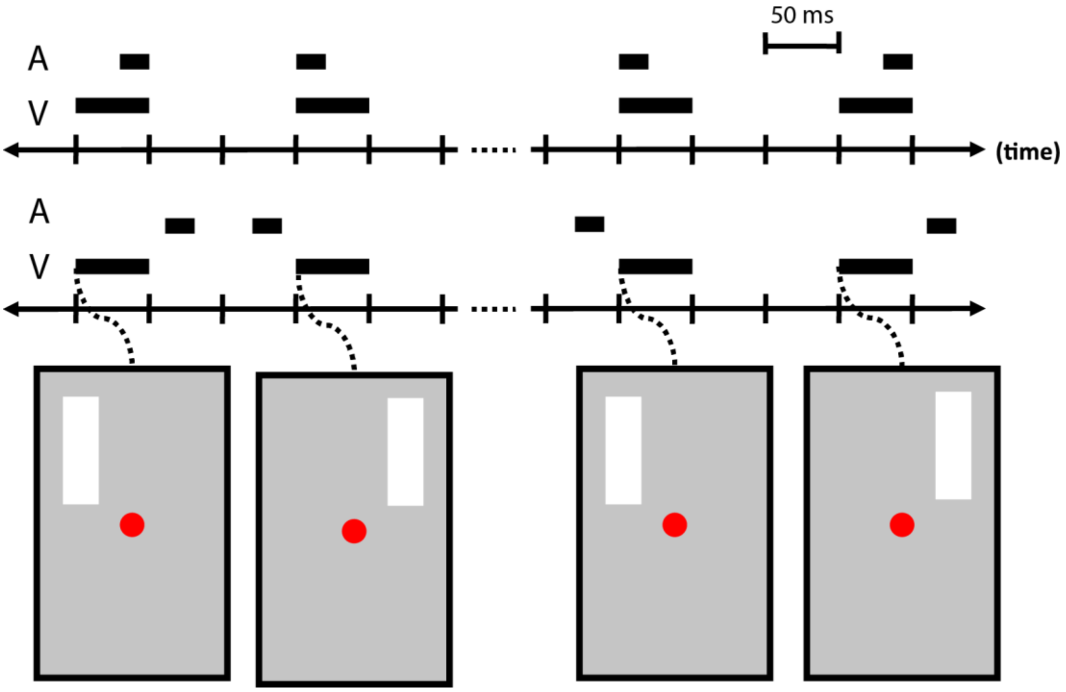


Fig. 37. Audiovisual conditions used in the visual time interval and apparent motion experiments. Demonstrations of the trials for mild (top) and extreme (bottom) auditory modulation conditions at the temporal dimension. Short (left) and long (right) auditory modulations are given in random order.

The bar is presented twice first either on the left or right randomly and in the second time presented on the contrary side. Boxes at bottom illustrate how the visual stimuli change in time.

5.2.4 *Data acquisition, preprocessing, ERP, and time-frequency analyses*

All the data acquisition, preprocessing, ERP and time-frequency analyses were the same as in the Experiment 1&2.

5.3 Results

5.3.1 *Amount of auditory modulation influences perception of apparent motion speed*

As Fig. 38 shows participants' perception of speed was significantly affected in the same direction of auditory clicks in both mild modulation condition (M = 73.3, SD: 10.83); $t_{18} = 9.37$, $P < 0.001$, but especially in extreme modulation comparison (M = 91.2, SD:5.62) condition; $t_{18} = 31.96$, $P < 0.001$. When SOA was extreme (45 ms), the percentage of perceived visual interval affected by auditory interval was significantly higher compared to the mild modulation case (SOA: 15 ms); $t_{18} = 8.47$, $P < 0.001$.

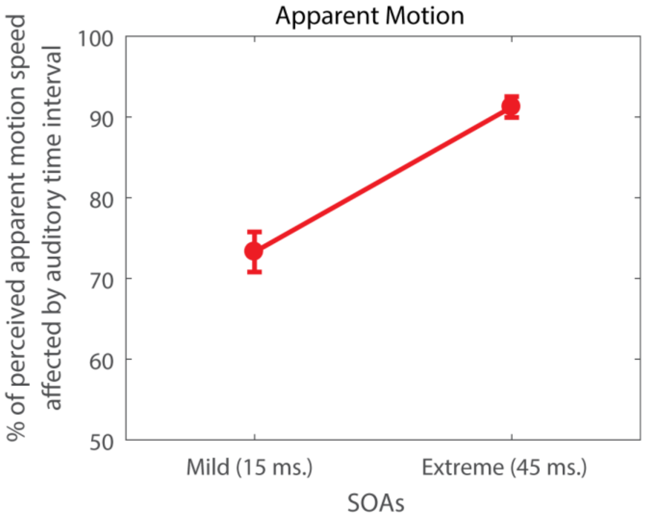


Fig. 38. Percentage values indicating the effects of different auditory time interval conditions. Extreme auditory modulation affected speed perception of visual stimuli more than mild modulation, also the effects were above the chance level of %50. Two levels of discrepancies between auditory time intervals were presented in the x-axis.

5.3.2 *Auditory modulation effects on ERPs elicited by visual apparent motion: the role of various regions and time intervals on distinct features*

We observed behavioral results that shows the auditory interference on the visual speed perception, and the level of perceptual change is positively correlated with the level of discrepancy. We excepted to see the effect of both auditory duration and the levels of the discrepancies. Therefore, we may consider two physiological hypotheses: (1) auditory time intervals could interact with the visual speed processing at a location specific to the speed processing at various degrees depending on the auditory temporal structure, in this case, the interaction would be an issue of the amount of excitation at that region: *localized interaction*. (2) However, another hypothesis is that interaction may occur at different locations at different time windows: *distributed interaction* and the overall speed perception is affected. Localized interaction hypothesis presumes a late interaction at a distinct region, whereas distributed interactions approach would assume interactions at both early and late steps of the speed perception.

As in Experiment 2, we wanted to determine how distinct auditory time intervals presented concurrently to apparent motion affect evoked activities elicited by visual apparent motion (i.e. the visual test stimulus which is the same across all conditions). To reveal the spatiotemporal profile of the interactions, repeated measures ANOVAs and *t*-tests were applied. As shown in Fig. 39A, there was a clear interaction effect within 150-180 ms, and as late as 480-540 ms mostly over central posterior regions. In addition, short and long auditory modulations, as the main effect, caused differences mostly at right central and parietal regions similarly over 140-200-ms time window (see also upper and lower panels in Fig. 40). Additionally, mild and extreme modulation effect differences were observed over 250-300-ms time window, especially over right frontal regions (see also Fig. 41, upper panels).

Further *t*-test comparisons of short and long auditory time interval modulations had different spatiotemporal profiles for mild and extreme levels of audiovisual discrepancies (Fig. 39B). Extreme short and extreme long auditory stimuli comparison showed an effect on visual processing in the late time window of 510-550 ms, mostly over medial centroparietal regions (see also Fig. 40, upper panel). On the other hand, differences in mild short and long auditory effect on visual processing was observed mostly in 150-190-ms time window, in the right centroparietal region (see also Fig. 41, lower panel); and frontal very early differences were also observed.

As for comparisons of mild and extreme modulations on visual processing for both short and long auditory intervals (Fig. 39C), they had similar temporal distribution in the right frontal where mild created higher potentials in 260-290-ms time window, observed at the main effect (see also Fig. 41, upper panels). However, differences in mild and extreme auditory influence were different for short and long discrepancies over the medial parietal region. Although there was no difference between mild and extreme modulations on visual processing for longer auditory stimuli (lower panel in Fig. 39C); for shorter auditory stimuli, a difference was observed at several time windows that can be observed in the upper panel of Fig. 39C.

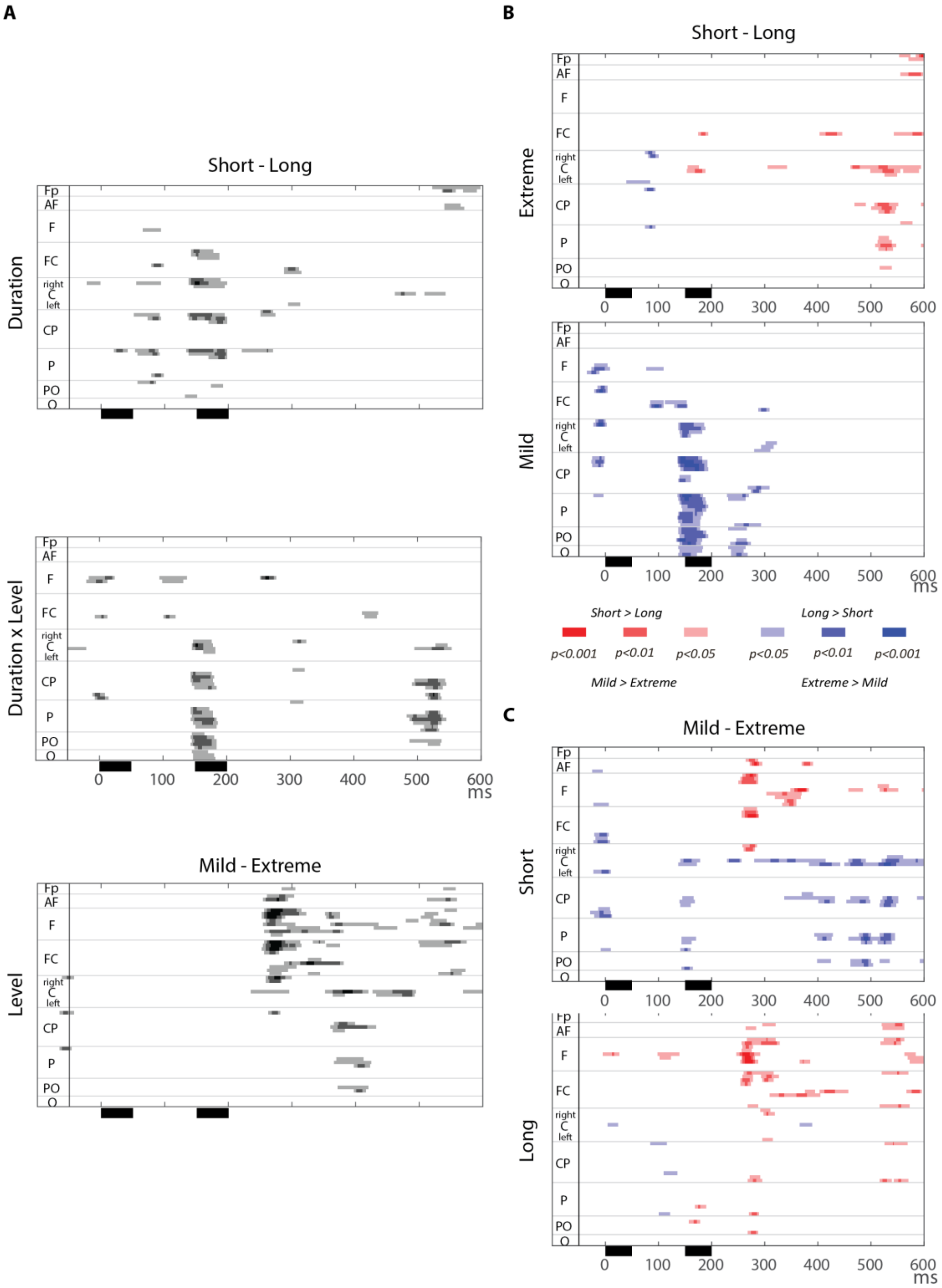


Fig. 39. Time courses of auditory effects on apparent motion evoked activity. Apparent motion frames are depicted by black rectangles at the bottom of the timeline. Other conventions are the same as those in Fig. 29.

Fig. 40. shows the average potential changes on the sites where the main interaction of the independent variables is observed. As this figure indicated, the difference in the mild discrepancy is processed mostly at the early time window in a way that shorter creates higher potential, however, in the high discrepancy conditions, the longer creates a higher potential. This result indicates a qualitative difference at the site and time window. Further comparison with t-tests indicates that this region is more sensitive to mild discrepancies at early time window, whereas it is also sensitive also to high discrepancies at a late time windows. Therefore, interactions in ERPs over this region suggest a selective role of the region in temporal processing. This result also supports distributed interactions hypothesis.

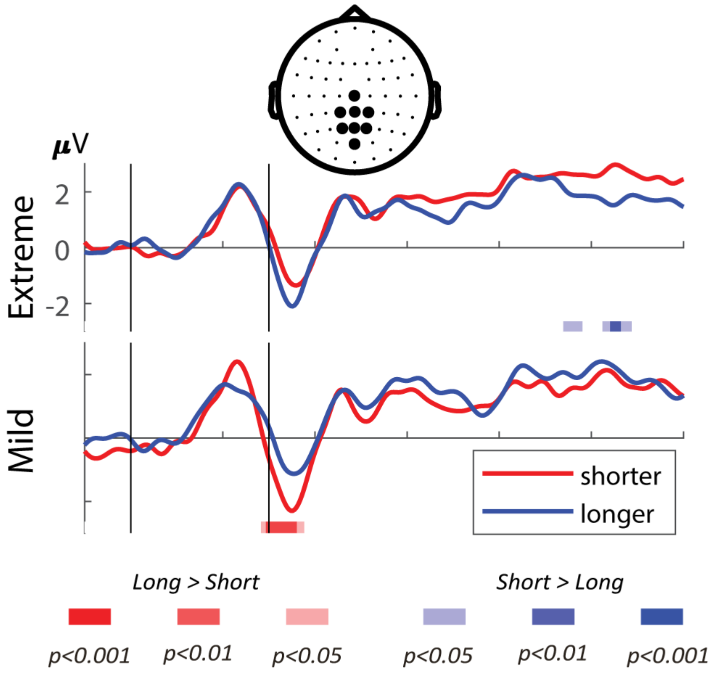


Fig. 40. **AV-A** ERPs comparisons for shorter and longer auditory modulations of visual apparent motion, separately in extreme (upper panel) and mild (lower panel) levels at medial parietal region (Cz, CP1, CPz, CP2, P1, Pz, P2, POz). Other conventions are the same as those in the Fig. 31.

Observations from spatiotemporal interaction maps identified the role of two more regions (see Fig. 41). Right parieto-temporal region showed a difference when shorter and longer auditory intervals were presented. T-test comparisons of shorter and longer AV-As indicated that these differences were stronger in mild discrepancy condition, whereas it was not significant for extreme discrepancies. Although the results are similar to the medial parietal region potentials for the extreme are in the reverse direction. Therefore, this site processes the extreme auditory time intervals differently. On the other hand, the right frontal site is sensitive to the amount of discrepancy, especially at 260-290-ms time window.

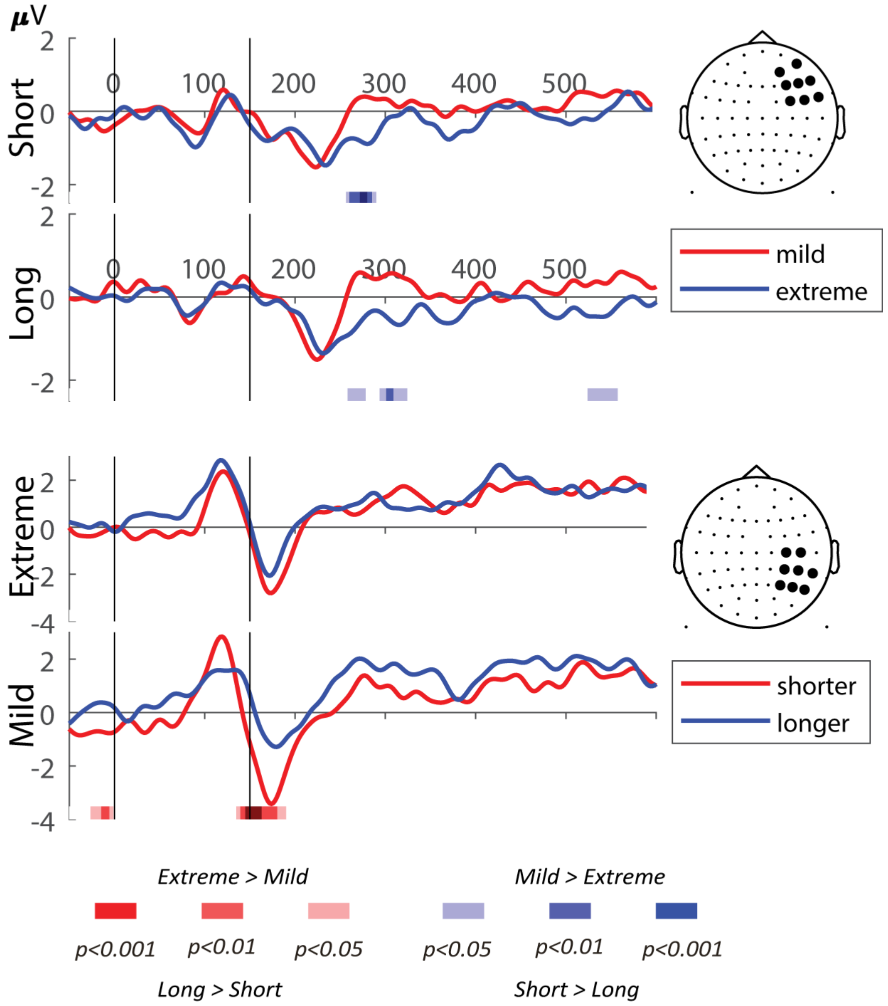


Fig. 41. **AV-A** ERPs comparisons for mild and extreme auditory discrepancies of visual apparent motion, separately in short (first panel) and long (second panel) auditory levels at right frontal region (AF4, AF8, F4, F6, F8, FC4, FC6, FT8). Comparisons for shorter and longer auditory modulations of visual apparent motion, separately in extreme (third panel) and mild (fourth panel) levels at right parietal region (C4, C6, CP4, CP6, TP8, P4, P6, P8). Other conventions are the same as those in the Fig. 31.

To sum up, as we attempted to decipher the auditory effect on visual speed perception, ERP analyses over AV-A responses revealed three distinct sites having a role in this process. First, right parietal region identifies the auditory effect on visual speed perception when shorter and longer auditory time intervals were compared, especially in case of mild discrepancy. Second, right frontal sites identify the amount of temporal discrepancy between auditory and visual presentation. Third, medial posterior sites discriminate the auditory effects between shorter and longer presentations of the auditory time intervals in different ways for mild and extreme discrepancies. These results suggest the distributed interactions of the auditory responses in the scope of this experiment.

5.3.3 *Oscillatory evoked-power of the frontal ROI changes as a function of the direction of the auditory modulation while occipital is to the amount of the SOA values.*

To understand the oscillatory dynamics of this experiment in detail, spectral amplitude and phase values were obtained through wavelet transform (see Fig. 42). Similar to the Experiment 2, the same sites show that frontal ROIs indicates a different power pattern than parietal sites, however auditory only stimuli showed more salient effects in this apparent motion experiment compared to the time interval judgment in Experiment 2. Shorter time interval modulation of auditory stimuli creates more power in both extreme and mild levels of SOA ($F_{1,12} = 13.902$, $P = 0.003$, $\eta_p^2 = 0.537$). Interestingly, this effect was observed at the AV-A responses, too ($F_{1,9} = 8.541$, $P = 0.017$, $\eta_p^2 = 0.487$). No effects of auditory duration or modulation level on the power durations of the responses were observed.

Spectral maps from right parietal region revealed that two frequency bands had high power components, 4-6 Hz, and 6-12 Hz (see Fig. 42, last column). Power and duration analyses were applied to these frequency bands. As revealed in Fig. 42, first column, mild interactions of shorter and longer auditory stimuli created more power for AV conditions in the occipital regions at 4-6 Hz band ($F_{1,16} = 17.614$, $P < 0.001$, $\eta_p^2 = 0.524$, see Fig. 43). In addition, parallel to the ERP results, spectral powers of the AV-A responses also revealed the auditory duration effect both in terms of the power amplitude ($F_{1,12} = 11.281$, $P < 0.01$, $\eta_p^2 = 0.485$, see Fig. 43, upper row last column) and power durations ($F_{1,12} = 5.452$, $P < 0.05$, $\eta_p^2 = 0.312$, see Fig. 43, lower row last column). However, auditory-only spectral maps showed no significant difference across conditions. Other regions frequency band did not indicate any interaction effect. To sum up, 4-6 Hz frequency maps of AV-A responses from the right parietal region is sensitive to the duration of the auditory stimulus in both power amplitudes and durations. Furthermore, mild modulation caused higher power in this region ROI in general, and this effect is more salient in shorter-interval auditory modulation compared to longer ones.

To reveal the spectral differences in right frontal sites that discriminate the differences in discrepancy, these sites were also analyzed. Mild short and long auditory-only simulation created more amplitude ($F_{1,13} = 5.076$, $P = 0.042$, $\eta_p^2 = 0.281$) and duration ($F_{1,13} = 5.642$, $P = 0.034$, $\eta_p^2 = 0.303$) of the spectral power (see Fig. 44). However, discrepancies that were observed in the auditory-only was not observed in AV and AV-A responses.

The sites that were sensitive to the interaction of the discrepancy level and auditory duration in ERP analysis also showed significant interaction in the power amplitudes at the 4-6-Hz band for both AV and AV-A responses ($F_{1,17} = 7.551$, $P = 0.014$, $\eta_p^2 = 0.308$; see Fig. 45, second row). Additionally, like the right parietal region, AV responses also revealed that mild discrepancies created more power also at this region ($F_{1,17} = 24.149$, $P = 0.000$, $\eta_p^2 = 0.587$).

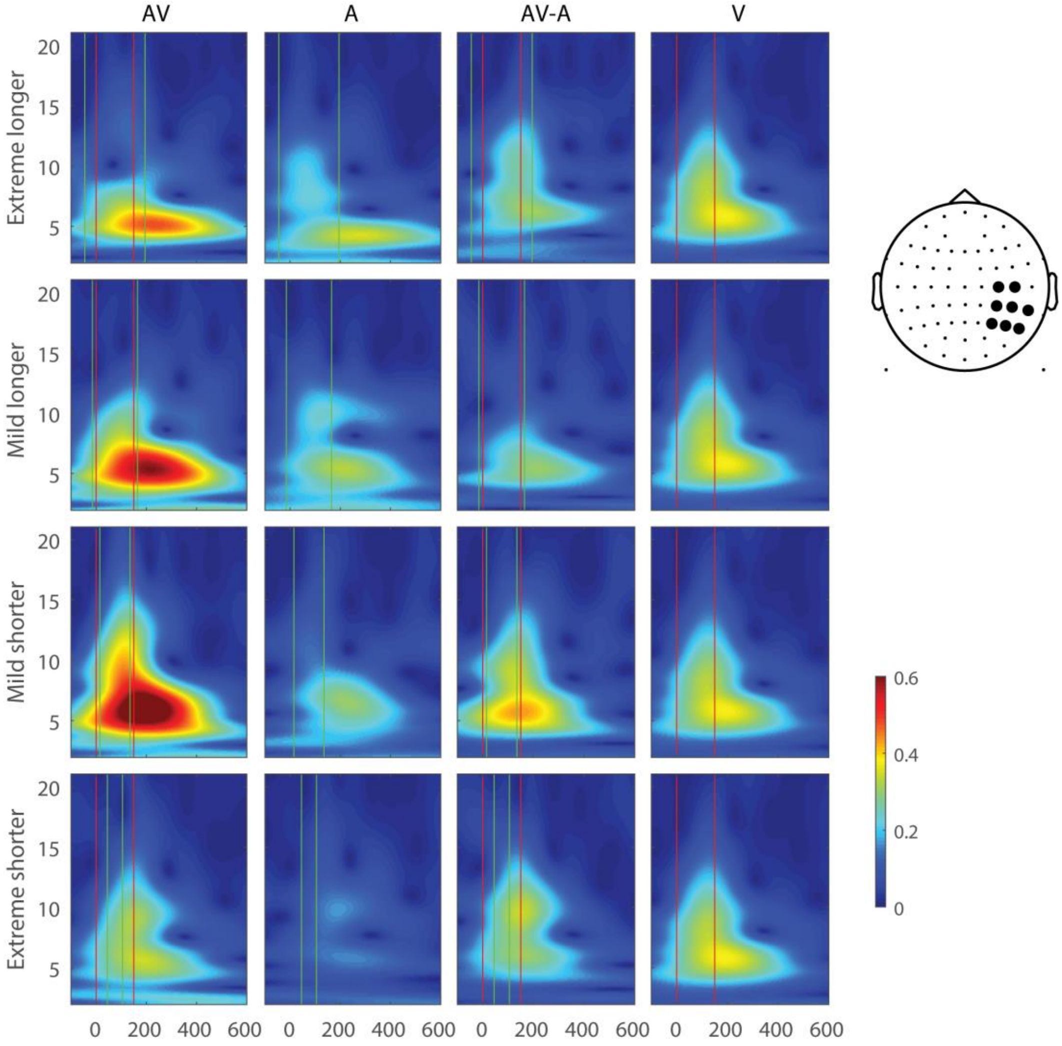


Fig. 42. Average signal from right parietal sites was averaged for calculating the time-frequency decompositions on average AV, A, V, and AV-A modalities. Other conventions are the same as those in

Fig. 32.

Oscillatory analyses results were two-fold. Although, comparisons of the same sites with the Experiment 2 indicated similar role to the same sites, auditory duration effect on the AV-A responses in the frontal region was stronger in the present experiment. Spectral powers in the 4-8-Hz band of the specific sites that are sensitive to the auditory influences in ERP results indicated the same effects for parietal regions. However, only the right frontal site did not show the discrepancy level effect at any frequency range.

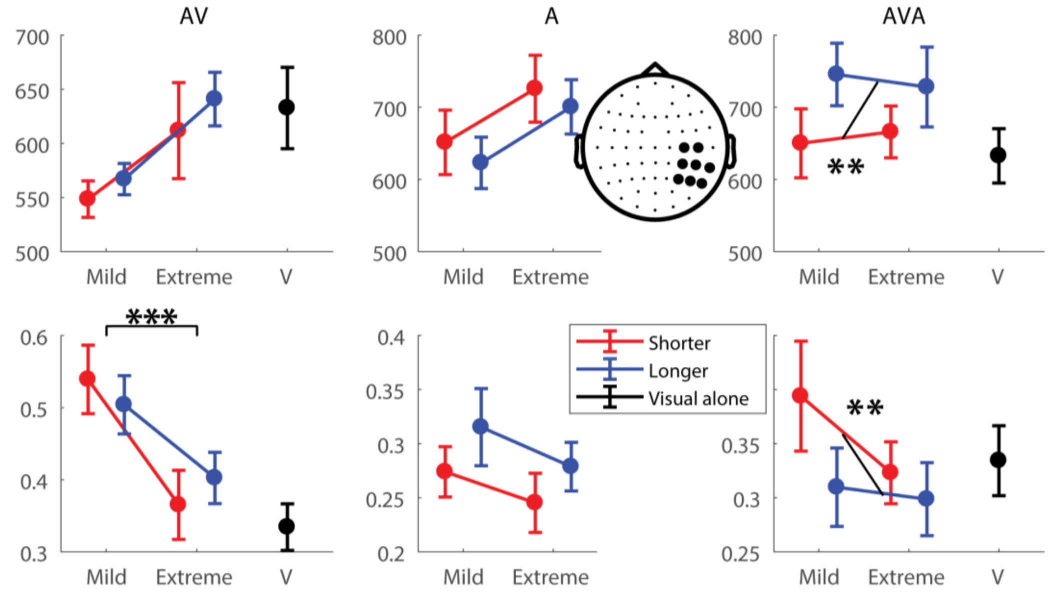


Fig. 43. Oscillatory power comparisons in the averaged left occipital channels (PO7, O1, Oz) for shorter and longer auditory modulations (frequency band: 4-8 Hz). (A) Average powers elicited by the shorter and longer auditory intervals. (B) Averaged oscillatory responses and duration of the responses for different modalities of auditory time interval conditions (bimodal: AV, unimodal: A, and synthetic AV-A) (*N* = 19). Other conventions are the same as those in Fig. 33.

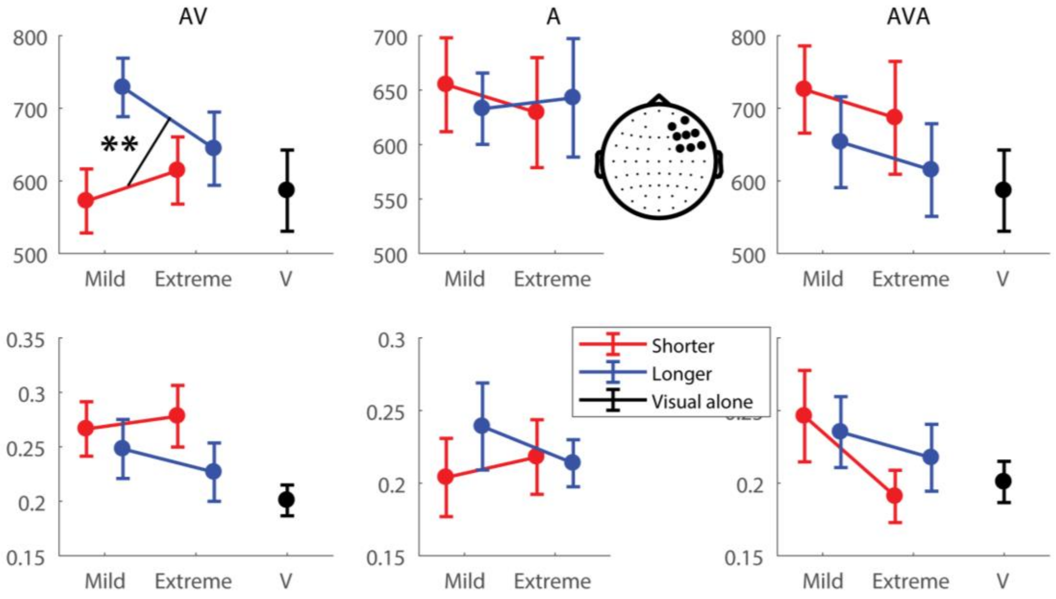


Fig. 44. Oscillatory power comparisons in the right frontal sites for shorter and longer auditory modulations (frequency band: 4-6 Hz). Other conventions are the same as those in Fig. 33.

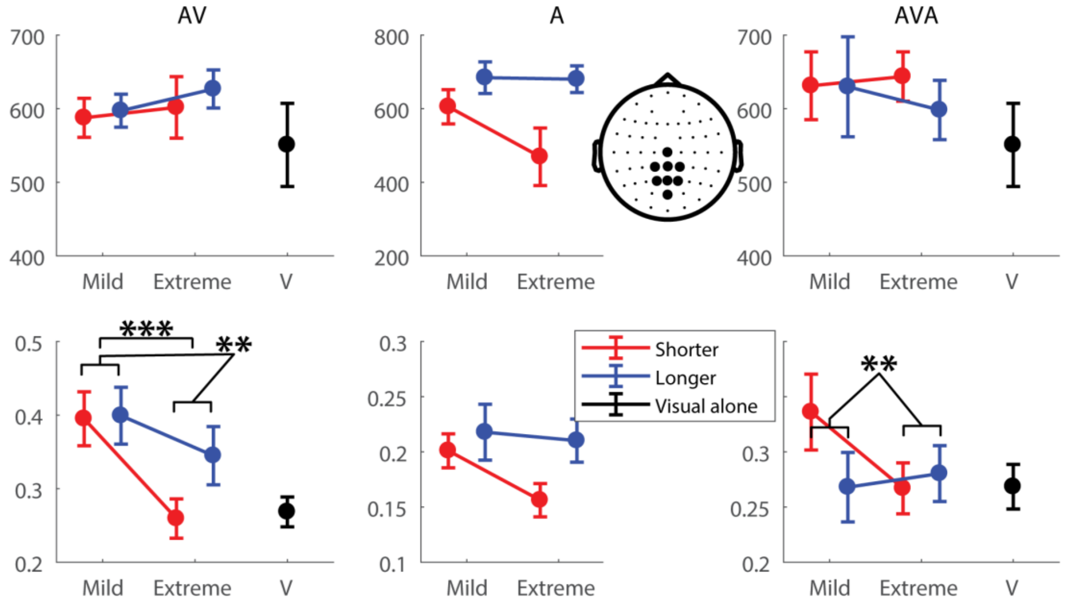


Fig. 45. Oscillatory power comparisons in the medial parietal sites for shorter and longer auditory modulations (frequency band: 4-6 Hz). Other conventions are the same as those in Fig. 33.

5.4 Discussions and Conclusion

In the present experiment, we systematically examined the evoked activities and spectral power to the visual apparent motion under different concurrent auditory time interval conditions. This experiment revealed several interesting and novel findings. Based on the time interval (shorter or longer than that of the apparent motion) and the level of the discrepancy of the auditory time intervals (mild or extreme), we found significant and distinct changes in the ERPs and spectral powers elicited by visual apparent motion. First, auditory interaction on the visual speed perception occurs as early as the interaction in the first frame of the apparent motion at the right parietal region. However, when effects of mild and extreme discrepancies considered, they showed differences at the same time interval at the medial parietal region, which might indicate different pathways for detecting the direction of the first frame of audiovisual stimulus and the amount of its discrepancy. Second, discrimination of the level of audiovisual discrepancies was observed at right frontal region after the second frame of the apparent motion. Finally, 4-8-Hz theta band spectral power had an important role in the auditory interaction.

5.4.1 Distributed interactions of audiovisual temporal integration of speed perception

Findings in the current experiment demonstrate that presentation of the concurrent shorter and longer time intervals significantly alters the ERPs elicited by visual apparent motion. By indicating specific ERP components and spectral powers at specific scalp sites, this experiment here shed light on the nature of the cortical processes involved in concurrent auditory time intervals on motion perception. It is well known that the temporal and spatial information provided by audition can affect visual motion perception (Soto-Faraco *et al.*, 2003; Hidaka *et al.*, 2015). Significant influences of auditory time interval on visual motion perception have been found in psychophysical experiments (Getzmann, 2007; Freeman & Driver, 2008; Kafaligonul & Stoner, 2010). These psychophysical studies suggest that the visual system actively uses auditory timing to estimate motion, the role of the auditory timing in shaping motion perception and the corresponding neural activities are still unclear. Although we aimed to decipher the auditory effects on visual processing, concurrent presentation of the auditory time intervals elicits both auditory and visual evoked responses simultaneously. However, removing auditory evoked components from audiovisual responses enabled us to compare remaining visual components. Our results indicate that the significant differences between AV-A ERPs and theta spectral powers for shorter and longer auditory time intervals are located over parietal electrodes. The amplitudes of early components (70- to 100-ms and 140- to 180- ms time ranges) are significantly affected. No further later auditory duration effects are observed at early(<200 ms) stages of the processing. These results confirm the recent findings emphasizes the importance of the right parietal region in sub-second duration judgment (Battelli *et al.*, 2007; Salvioni *et al.*, 2013; Hussain Shuler, 2016; Kaya *et al.*, 2017). Furthermore, sub-second duration judgments were also affected by the application of brain stimulation to right parietal cortices (Bueti *et al.*, 2008; Dormal *et al.*, 2016). In addition to the role of this region, the medial parietal region also indicated different interaction patterns in discriminating shorter and longer auditory effects between mild and extreme discrimination conditions. Therefore, these results support the idea that features of the auditory time intervals (being short or long or having low and high discrepancy) affects the visual processing at different pathways and different time windows. Nevertheless, in this apparent motion experiment, we can conclude that short and long auditory effects are present at right parietal region before the second frame of the apparent motion with no effect of discrepancy and right frontal region discriminated the level discrepancies with no effect of duration. Effects of the discrepancy are observed at later (>250 ms) processing steps.

It is important to notice that interactions of these early processing steps correspond to the evoked potentials elicited by the first frame of the apparent motion. Therefore, it is not possible to relate the differences to a complete interval. Although the physical stimuli are the same, ERPs evoked by the first frame in Experiment 2 are different from the present experiment. These findings suggest that participants' task-related expectancies affect the way how the auditory information play a role in visual speed processing.

Time intervals provided by auditory stimulation significantly affected the perceived speed of apparent motion. Participants compared physically the same visual apparent motions however auditory stimulus with either shorter or longer time intervals concurrently presented with visual stimuli. As a result, participants perceived the visual apparent motions faster, when the shorter auditory interval was presented; and, perceived slower, when longer auditory interval presented. The percentages of the speed perception affected by the same direction of the auditory stimulus were higher when the audiovisual discrepancy is high. This quantitative difference in perceptual measurement was supported difference of the mild and extreme discrepancy ERPs over the frontal region. In that region, additional to the short and long differences extreme discrepancies were also discriminated.

Audition has an important role to identify the motion on the visual field and estimate its parameters. Compare to visual time interval perception in Experiment 2, speed perceptions have an extra survival function in visual cognition and auditory information may have both alarming and speed accuracy boosting role in visual processing. Compare to the Experiment 2, the levels of the significances in the present experiment were higher, and this is compatible with the behavioral results. Interestingly in occipital area, it is easy to see the differences caused by short and long clicks on each rising edges of the peaks.

In addition to the results in Experiment 2, time-frequency decompositions of the different modalities of the experimental conditions revealed that auditory only stimulus created significantly different powers between mild and extreme conditions, even though the auditory stimuli were the same, and the task was totally visual oriented. This can be interpreted that participants were more sensitive to the auditory stimulation due to its importance in the judgment of the speed of visual motion.

Behavioral results of the present experiment indicate that the size of discrepancy effects the amount of the apparent motions whose speed were perceived in the same direction of the auditory time interval. Quantitative change in stimulus results a quantitative difference in perception may suggest a role for a common timing mechanism for the discrepancy processing. As opposed to this view, we found distinct spatiotemporal characteristics of auditory time interval interactions. Accordingly, our ERP and spectral power results emphasize the recruitment of distinct regions rather than an amplitude modulation in a specific location in visual motion perception.

5.5 Limitation and future directions

Differences in the interactions of the first frames were different in the first, second and the third experiments. This indicates an effect of the pre-stimulus state of the ongoing activities. To investigate the pre-stimulus activities, an experiment could be designed. Also in order to judge the role of auditory interaction in the beginning and the end of the motion, SOAs of the first and the second frames can be modulated separately and the behavioral and consequently the effect can be observed. Considering the task of the second and the third experiments, due to possible entrainment of the working memory loads, first and the second time intervals and apparent motions can be investigated separately.

CHAPTER 6

6 EXPERIMENT 4. THE INVOLVEMENT OF CENTRALIZED AND DISTRIBUTED PROCESSES IN SUB-SECOND TIME INTERVAL ADAPTATION: AN ERP INVESTIGATION OF APPARENT MOTION [1]

ABSTRACT

Accumulating evidence suggests that the timing of brief stationary sounds affects visual motion perception. Recent studies have shown that auditory time interval can alter apparent motion perception not only through concurrent stimulation but also through brief adaptation. The adaptation aftereffects for auditory time intervals were found to be similar to those for visual time intervals, suggesting the involvement of a central timing mechanism. To understand the nature of cortical processes underlying such aftereffects, we adapted observers to different time intervals by using either brief sounds or visual flashes and examined the evoked activity to the subsequently presented visual apparent motion. Both auditory and visual time interval adaptation led to significant changes in the ERPs elicited by the apparent motion. However, these changes were in the opposite direction. Also, for each modality, they mainly occurred in different time windows and clustered over distinct scalp sites. The effects of auditory time interval adaptation were centered over parietal and parieto-central electrodes while the visual adaptation effects were mostly over occipital and parieto-occipital regions. Moreover, the changes were much more salient when sounds were used during the adaptation phase. Taken together, our findings within the context of visual motion point to auditory dominance in the temporal domain and highlight the distinct nature of the sensory processes involved in auditory and visual time interval adaptation.

Keywords: time interval, sub-second adaptation, auditory, visual, apparent motion, event-related potentials

---

[1] This chapter is adapted from Kaya *et al.* (2017)

6.1 Introduction

Time perception is essential for survival in a dynamic world. To successfully interact with the external environment, the brain processes temporal information over a wide range of timescales from milliseconds up to several hours. A fundamental aspect of time perception is the ability to demarcate temporal intervals. To date, many different views and models have been proposed to account for time interval perception (Mauk & Buonomano, 2004; Grondin, 2010; Merchant *et al.*, 2013). Certain models posit a central common mechanism (e.g., a supramodal clock) to explain time interval perception and the corresponding neural activity. Other models asserted multiple distributed mechanisms and context dependent (e.g., modality-specific) processes. Several studies also suggested the existence of both a central mechanism activated by a variety of timing tasks and the neural structures that are selectively engaged by different timing contexts. Accordingly, timing models have been developed to include both central and distributed neural mechanisms by proposing that time interval estimation depends on the interaction between these processes (Merchant & de Lafuente, 2014).

To understand the contribution of central and distributed processes to the perception of sub-second time intervals, audiovisual paradigms have been extensively used. The findings along this direction have highlighted the distinct temporal characteristics of the two modalities. In general, sensitivity to the time intervals and rhythms marked by the auditory signals are better than to those by visual signals (Welch *et al.*, 1986; Burr *et al.*, 2009; Grahn *et al.*, 2011; Rammsayer *et al.*, 2015). When both signals are introduced together, the timing of brief sounds (i.e., auditory time intervals) drives the perceived timing of visual events (Fendrich & Corballis, 2001; Morein-Zamir *et al.*, 2003; Recanzone, 2003). This phenomenon is called temporal ventriloquism and is generally interpreted as a perceptual consequence of the superior temporal resolution of the auditory system. More importantly, it has been shown that crossmodal temporal adaptation occurs between audition and vision, such that being adapted to a temporal property in one modality affects the perception of that specific property introduced through the other one (Murai *et al.*, 2016). For instance, adapting observers to a certain auditory rate has significantly changed the perceived rate of subsequent visual flickers, and vice versa (Levitan *et al.*, 2015). These results point to sensory processes operating over two modalities and hence point to the involvement of central timing mechanisms. On the other hand, Li *et al.* (2015) have provided evidence that duration adaptation aftereffects for audition and vision can be obtained concurrently, suggesting that modality-specific duration adaptation can operate in parallel. They also found that these concurrent aftereffects of audition and vision are contingent on low-level auditory but not visual attributes. Accordingly, these findings emphasize the recruitment of different processes in the aftereffects induced by each modality (see also Bruno & Cicchini, 2016).

Auditory dominance in time percepts has been illustrated in other aspects of vision which do not require explicit timing judgments, but which significantly depend on the timing of visual events. Several studies found that the timing of auditory signals (i.e., the time interval defined by brief static sounds) can alter the perception of visual apparent motion, presumably by altering the perceived timing of the individual frames of the apparent motion stimuli and/or the interstimulus intervals (ISIs) between those frames (Getzmann, 2007; Freeman & Driver, 2008; Kafaligonul & Stoner, 2010, 2012; Shi *et al.*, 2010). Moreover, using a temporal adaptation paradigm, Zhang et al. (2012) investigated the influence of time interval adaptation (auditory or visual) on motion perception. Their results showed that adapting observers to different time intervals induced significant aftereffects on the perception of visual Ternus display (i.e., apparent motion). Adaptation to auditory and visual time intervals induced similar aftereffects, highlighting the role of a central supramodal mechanism in the observed aftereffects for motion perception.

These findings are interesting by suggesting that not only concurrent presentation of auditory time intervals but also a previous experience acquired through a brief adaptation phase is involved in shaping visual motion perception. However, the neural mechanisms underlying these adaptation aftereffects remain unclear. A fundamental question is whether they are mediated mostly by a central supramodal mechanism or by distributed (modality-specific and/or crossmodal) sensory processes. In the current experiment, we aimed to address this question by looking for the neural correlates of these aftereffects on visual motion processing. We adapted observers to different time intervals by using brief static sounds or visual flashes and recorded EEG activity for the subsequently presented visual apparent motion. To identify the contribution of central and distributed processes, we investigated the common and distinct effects of auditory and visual time interval adaptations on the ERPs across all scalp sites. We predicted that a central supramodal process would be revealed by common electrode sites and ERP components at which the aftereffects become significant for both modalities. On the other hand, any distinct aftereffects of audition and vision on the spatiotemporal profile of the neural activity would imply the involvement of specific processes for each modality. In addition to testing this basic hypothesis on sub-second time interval adaptation, by identifying scalp sites at which the auditory time interval aftereffects take place, we were also able to shed light on how the temporal information provided by audition is used by the sensory system for visual motion processing.

6.2 Materials and Methods

6.2.1 Participants

Eighteen observers (9 females; age range: 21-32 years) participated in this experiment with seventeen being naive to the purpose of the experiments. 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 Ankara University.

6.2.2 Apparatus and Stimuli

We used Matlab version 7.12 (The MathWorks, Natick, MA) with the Psychtoolbox 3.0 for stimulus presentation and data acquisition. Visual stimuli were presented on a 20-inch CRT monitor (Mitsubishi Diamond Pro 2070SB, 1280 × 1024 pixel resolution and 100 Hz refresh rate) at a viewing distance of 57 cm. A SpectroCAL (Cambridge Research Systems, Rochester, Kent, UK) photometer was used for luminance calibration and gamma correction of the display. Sounds were presented via insert earphones (EARTone 3A, Etymotic Research, Village, IL) and amplitudes were measured by a sound-level meter (SL-4010, Lutron Electronics, Taipei, TW). The timing of auditory and visual stimuli was confirmed with a digital oscilloscope (Rigol DS 10204B, GmbH, Puchheim, Germany) connected to the computer soundcard and a photodiode (which detected visual stimulus onsets). All experiments were performed in a silent, dimly lit room.

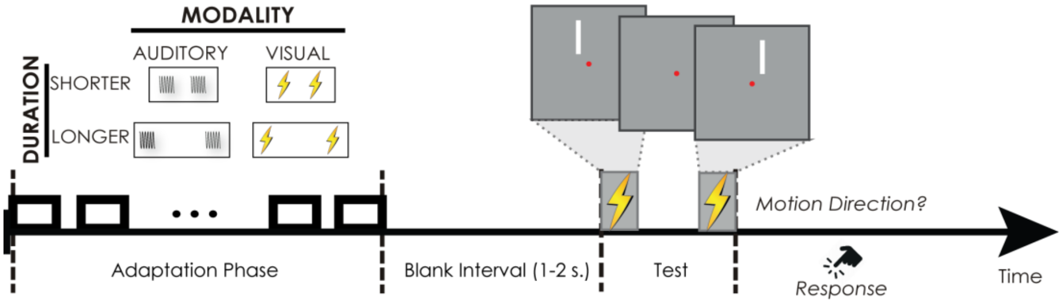


Fig. 46. Experiment design. Each trial consisted of an adaptation and a test phase. During the adaptation phase, either two auditory clicks or flashed bars were presented sequentially. The time interval of these adaptors (two auditory clicks or two visual flashes) was either shorter or longer than the ISI between the motion frames. After a variable blank interval, two-frame apparent motion (test stimulus) was displayed and observers were asked to report the direction of motion. Each adaptation condition (2 modalities × 2 time intervals) was run in separate experimental blocks. The test stimulus was a two-frame apparent motion. The time interval (ISI) between each motion frame (i.e. left and right visual bars) was fixed across different adaptation conditions.

A small red circle (0.3 deg diameter) at the center of the display served as a fixation target. We used two-frame apparent motion (AM) as a test stimulus. The apparent motion consisted of two “flashed” (50 ms) bars (0.4 × 3.0 deg) centered 3 deg above the fixation circle (Fig. 46). The spatial displacement (i.e. center-to-center separation) and the inter-stimulus interval (ISI) between each flashed bar were 1 deg and 100 ms, respectively. Each bar was brighter (107 cd/m2) than the grey background (21 cd/m2). Adapting stimuli were either visual or auditory (Fig. 46). To adapt observers to visual time intervals, we used the same visual bars used for the apparent motion. However, instead of having a spatial separation between them, each visual bar was flashed (50 ms) at the same location (3 deg above the fixation circle). The time interval (ISI) between the flashed bars was either 50 ms (shorter time interval condition) or 350 ms (longer time interval condition). For auditory time interval adaptation, two successive static clicks (50-ms duration) were used. Each click comprised of a rectangular windowed 480 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization and was binaurally introduced at 75 dB sound pressure level (SPL). As in the visual adaptation conditions, the ISI values for shorter and longer conditions were 50 and 350 ms, respectively.

6.2.3 Task and Procedure

Each trial consisted of an adaptation and a test phase (Fig. 46). During the adaptation phase, either two auditory clicks or two flashed bars with a fixed ISI were presented sequentially. The time interval between each consecutive presentation was 600 ms, which was good enough to separate each repetition clearly for all adaptation conditions. At the beginning (first trial) of each experimental block, the adaptation phase consisted of 30 repetitions of successive auditory clicks or flashed bars. For the remainder of the trials, the adaptation phase included 8 repetitions. After a variable blank interval (1-2 sec), during which only the fixation point was present, the test pattern (apparent motion) was displayed. The direction of motion (rightward or leftward) was randomized across trials. At the end of each trial, observers were requested to indicate, by a key press, whether the test pattern moved rightward or leftward. Observers typically responded in less than 1 sec. When the response time was shorter than 1 sec, to ensure a sufficient inter-trial-interval, the next trial started 1 sec after the offset of the apparent motion. In case the participant failed to respond within 1 sec, the trial was discarded and repeated later. Observers were instructed to fixate throughout the trial and were also asked to pay attention to both the adapting and test stimuli.

Each adaptation (2 modalities x 2 time intervals) condition was run in a separate experimental block. In addition, observers completed an experimental block including only the test apparent motion (no adaptation condition). Instead of having an adaptation phase, each trial of this block started only with a fixation circle presented for a variable (1-3 sec) duration. Then, the test pattern was shown and the response was recorded as in the adaptation conditions. Each experimental block consisted of 40 trials in which the response times were shorter than 1 sec. The order of these five (4 adaptation and 1 no adaptation condition) experimental blocks were randomized across observers. Observers took a short (around 5 minutes) break between each block. Prior to the experimental blocks, each observer was shown examples of the test apparent motion, auditory and visual adapting stimuli.

6.2.4 EEG Data Acquisition

Electroencephalography (EEG) data were recorded with a 64 channel MR-compatible system (Brain Products, GmbH, Gilching, Germany), using sintered Ag/AgCl passive electrodes mounted on an elastic cap (BrainCap MR, Brain Products, GmbH, Gilching, Germany). The EEG caps included 63 scalp electrodes and an additional electrocardiogram (ECG) electrode attached to the back of each subject to control cardioballistic artifacts. The placement of scalp electrodes was based on the international 10/20 system. Two of the scalp electrodes, FCz and AFz, was the reference and ground electrodes, respectively. A syringe and q-tips were used to apply conductive paste (ABRALYT 2000, FMS, Herrschings–Breitbrunn, Germany) and to reduce impedances in each EEG channel. Electrode impedances were kept below 20 kΩ (typically below 10 kΩ) and they were monitored throughout the experiment for reliable recording. EEG signals were digitized at a 5 kHz sampling rate, and band-pass filtered between 0.016 and 250 Hz. Vision Recorder Software (BrainProducts, GmbH, Gilching, Germany) was used to store stimulus markers and EEG data on a secure hard disk for further analyses.

6.2.5 EEG Preprocessing and Analysis

We analyzed EEG data with BrainVision Analyzer (BrainProducts, GmbH, Gilching, Germany), the Fieldtrip toolbox (Oostenveld *et al.*, 2011) and our own Matlab scripts (The MathWorks, Natick, MA). For preprocessing, EEG signals were first down-sampled to 500 Hz and the cardioballistic artifacts were removed by using the signal from ECG channel. Then, the data were offline re-referenced to a common average reference and filtered through a zero phase shift Butterworth band-pass filter (0.5-70 Hz, 12 dB/octave) and a 50 Hz notch filter (50 Hz +/-2.5 Hz, 16th order). After the filtering process, the data was segmented into epochs -600 ms (before the onset of apparent motion) to 1200 ms (after the onset of apparent motion). An infomax independent component analysis was applied on these epochs to remove common EEG artifacts (e.g., eye blinks). Each trial was also screened automatically by artifact rejection criteria and manually by eye. In this semi-automatic artifact rejection, any trial with oscillations over 50 μV/ms or a voltage change more than 200 μV was rejected. On average 87.4% of trials were retained per condition.

After preprocessing, EEG signals from each specific electrode were averaged across trials to compute event-related potentials (ERPs) time-locked to the onset of visual apparent motion. A low-pass filter (40 Hz cut-off frequency) was applied to further smooth the ERPs. Baselines were also computed from a −100 to 0 ms time interval prior to the onset of the apparent motion and they were subtracted from the ERPs of each condition. During the adaptation phase of our experiments, the auditory clicks were introduced through headphones whereas the visual flashes were presented at the center of the apparent motion path. In other words, auditory clicks and visual flashes may act as distractors and as spatial cues, respectively. Therefore, more attention may be allocated to the apparent motion subsequently presented after the visual adaptation phase than the one after the auditory phase. Moreover, crossmodal and intramodal cueing has been found to be effective even at the earliest stages of visual cortical processing (Busse *et al.*, 2005; Ding *et al.*, 2014; Feng *et al.*, 2014). To overcome these and other (e.g., stimulus intensity differences) confounding factors due to the stimulation differences between two modalities, we initially restricted our ERP analyses to pairwise comparisons of time interval conditions for each modality separately. To determine specific electrode sites and components significantly affected by time interval adaptation, ERPs were compared between shorter and longer adaptation conditions using point-wise running t-tests (two-tailed). A significant difference in conditions was defined when at least 30 ms of contiguous data (i.e., 15 consecutive points at a sample rate of 500 Hz) met a 0.05 alpha criterion (for the validity of this approach see Guthrie & Buchwald, 1991; Senkowski *et al.*, 2007; Vroomen & Stekelenburg, 2010). In addition, over the identified scalp sites and time windows (TWs) based on both the running t-test results and difference (shorter-longer) topographies, we estimated the average ERP magnitudes for each subject. The significance of a consistent change in the average magnitudes was determined by using two-way repeated measures ANOVAs, with the modality (auditory and visual) and time interval (shorter and longer) as factors, and follow-up pairwise comparisons between shorter and longer time interval conditions.

6.2.6 Behavioral Data Analysis

For behavioral data, we only used trials included in the EEG analyses. In other words, the trials excluded during the EEG preprocessing stage were not considered for the analysis of the behavioral data. We calculated average performance and response time (RT) across subjects for each stimulus condition. To determine whether the effects modality of and/or time-interval are significant, we applied two-way repeated measures ANOVAs and follow-up pairwise comparisons to these behavioral measures.

6.3 Results

6.3.1 Behavioral Results

Fig. 47 shows the average performance on the motion discrimination for each condition. As expected, the motion direction discrimination performance was well above threshold and all observers reported the direction of apparent motion with more than 95% accuracy for all conditions. A two-way repeated-measures ANOVA on the accuracy scores, with modality (auditory and visual) and adaptation interval (shorter and longer) as factors, revealed a marginally significant modality effect (F1,17 = 4.250, P = 0.054, math formula = 0.200). The effect of time interval (F1,17 = 1.062, P = 0.317, math formula = 0.059) and its interaction with modality (F1,17 = 2.693, P = 0.119, math formula = 0.137) were not significant. The planned contrasts revealed only a marginally significant difference between the shorter and longer conditions of auditory adaptation (auditory t17 = −2.029, P = 0.068, visual t17 = −1.026, P = 0.804). None of the adaptation conditions were significantly different than the no adaptation condition (Bonferroni-corrected paired t-tests).

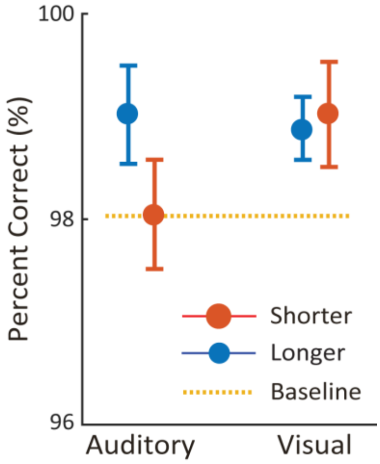


Fig. 47. Behavioral results on motion direction discrimination. Mean percentage correct values averaged across subjects (*N* = 18). The open and filled symbols correspond to shorter and longer time interval conditions, respectively. Error bars correspond to ± SEM. The dotted line indicates the mean value for the no adaptation condition, and the error bar placed over the symbol at the end of this line represents ± SEM.

6.3.2 Time interval aftereffects on the ERPs elicited by apparent motion: Time-course and scalp topographies

We wanted to determine whether adaptation to distinct time intervals could affect evoked activity to visual apparent motion (i.e., the test stimulus which is exactly the same across all conditions). If so, we should observe significant differences between the ERPs for shorter and longer interval adaptation conditions. Moreover, we reasoned that if the observed differences are due to a central (supramodal) timing mechanism, the changes in ERPs for both modalities should occur around similar times and electrode locations. Otherwise, we should observe specific changes in auditory and visual adaptation conditions. To test these possibilities, we performed pairwise comparisons (pointwise running t-tests with basic corrective steps) between shorter and longer ERPs across all individual electrode locations for each modality (auditory or visual) separately. Comparison of shorter and longer adaptation conditions of each modality separately also allowed us to limit potential confounding factors and to focus on adaptation aftereffects specific to time intervals (see Methods). For both modalities, the pairwise comparisons revealed significant differences between shorter and longer time interval conditions. However, these aftereffects were in the opposite direction. As shown in Fig. 48, we identified five different time windows in which the adaptation aftereffects take place. Except for a very early component (C1), the effects induced by auditory and visual time intervals occurred in different time windows and mostly clustered over distinct sites (Fig. 49).

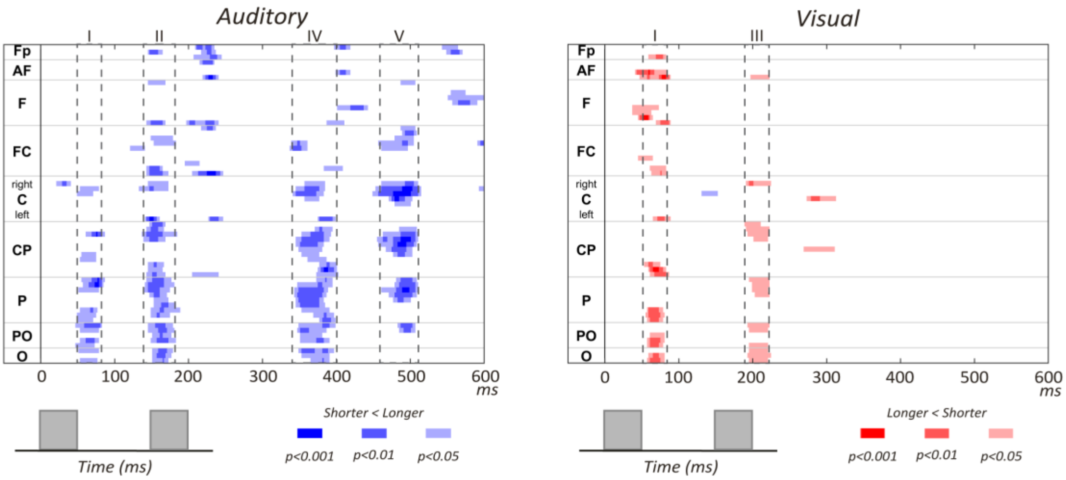


Fig. 48. Point-wise running t-tests on the evoked activity to the apparent motion presented after auditory and visual adaptors. The pairwise (shorter vs. longer) comparisons for each modality are shown in separate plots (auditory: left panel, visual: right panel). In each plot, time is displayed on the abscissa from 0 to 600 ms (relative to the apparent motion onset), and electrodes are displayed on the ordinate. The significant difference between time interval conditions is marked on the temporal axis by different levels. A specific time point was shaded only if at least 10 consecutive points were significantly different (see Methods). Based on the sign of the difference between shorter and longer time interval conditions, either blue (shorter < longer) or red (shorter > longer) color was used for shading. The identified time windows for further analysis are shown by the dashed rectangles (I: 50-80 ms, II: 140-180; III: 190-220 ms; IV: 340-400 ms; V: 460-510 ms). Apparent motion frames are represented by gray rectangles at the bottom of each plot. The timing and duration of apparent motion frames are indicated by the position and thickness of these rectangles, respectively.

In particular, the difference between the two time interval conditions was much more salient when auditory clicks were used during the adaptation phase. These differences were present within four different time windows. The activity within 50-80 ms was negative and positive over parieto-occipital electrodes for the shorter and longer auditory time intervals, respectively. This significant difference was present in both hemispheres and spread over neighboring occipital and parietal sites (Fig. 49A). Since the presentation of the second visual frame defines the time interval and apparent motion, the significant changes in ERP components after the onset of the second flash (150 ms) are particularly important in the current study. For the negative activity within 140-180 ms (N1 component), the ERPs for the shorter auditory time interval was significantly larger in amplitude than those for the longer one. This significant effect of time interval was centered at parietal regions and also existed at neighboring parieto-occipital and parieto-central electrodes. In addition, the effects of time interval adaptation within this time range were more salient in the right hemisphere, but also present in the left hemisphere. For auditory adaptation, significant differences between the two time interval conditions were also found for the components beyond 300 ms. The activity for the shorter time interval was less positive compared to the longer interval. As indicated by the topographic maps in Fig. 49A, the differential effects of adaptor time interval (shorter vs. longer) were centered over the parietal electrodes (and extended over the occipital and central regions) for the 340-390 ms range. The center of the cluster was medial relative to significant electrode locations for the 140-180 ms time window. For the 460-510 ms range, the significant differences were observed over centero-parietal sites in the right hemisphere.

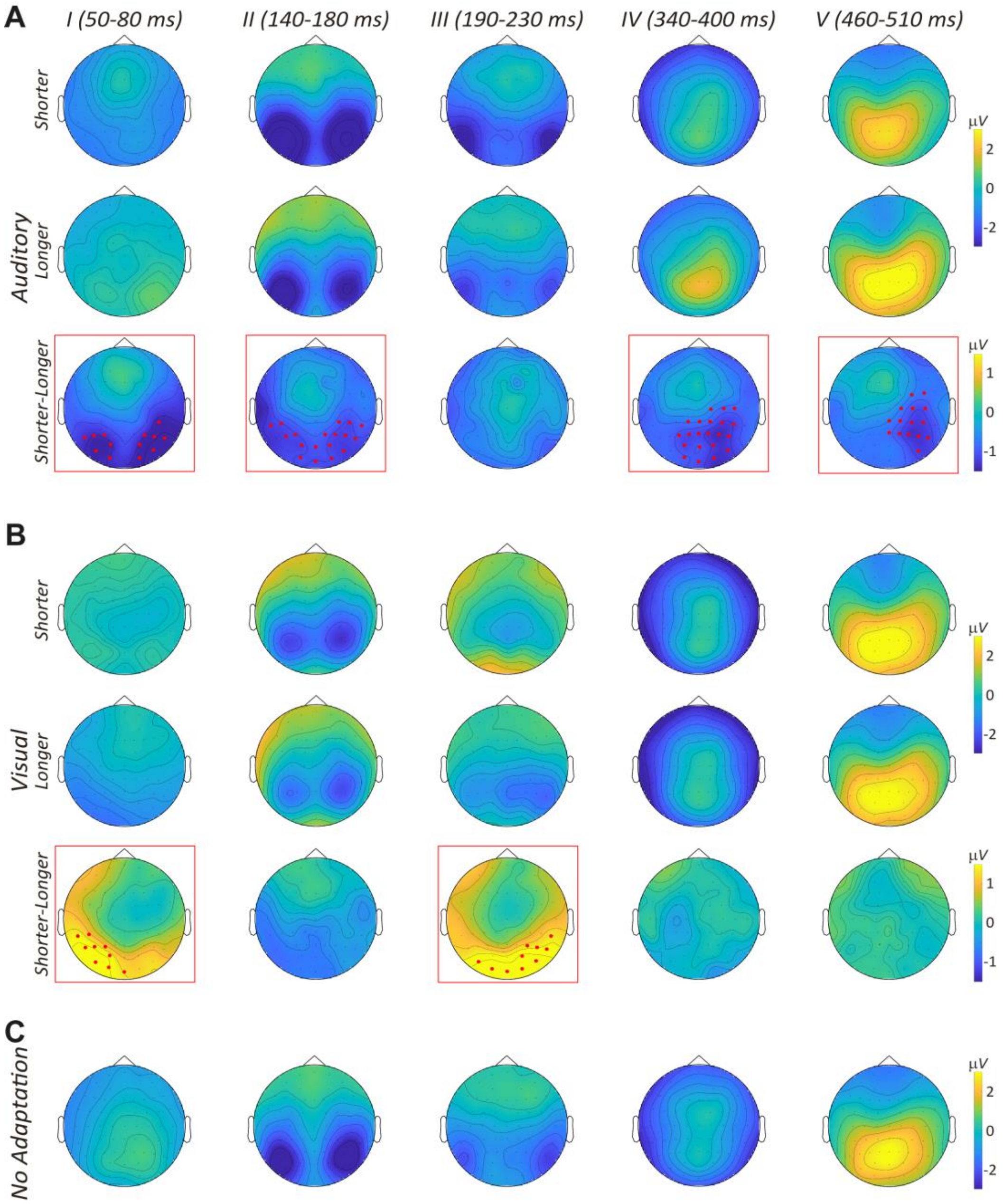


Fig. 49. Scalp topographies. Voltage topographical maps of the grand-averaged ERPs within each identified time window are displayed in separate columns. Each adaptation condition and the difference maps (shorter-longer) are shown in separate rows. The significant scalp sites (see also

Methods for additional criteria for significance) are marked by filled circles on the difference maps. (A) Auditory conditions, (B) Visual conditions, (C) No adaptation condition.

Compared to the differences between auditory shorter and longer conditions, the effects of visual time interval adaptation were less salient and occurred only in two time windows. We mainly observed differences between the two visual conditions in the 50-80 ms and the 190-220 ms time windows. The activity of the shorter condition was more negative than that of the longer condition within 50-80 ms. The aftereffects were clustered over parieto-occipital and parietal electrodes in the left hemisphere. In the 190-220 ms time range, over the occipital and parieto-occipital electrodes, the activity for the shorter condition was more positive than that for the longer condition (Fig. 49B).

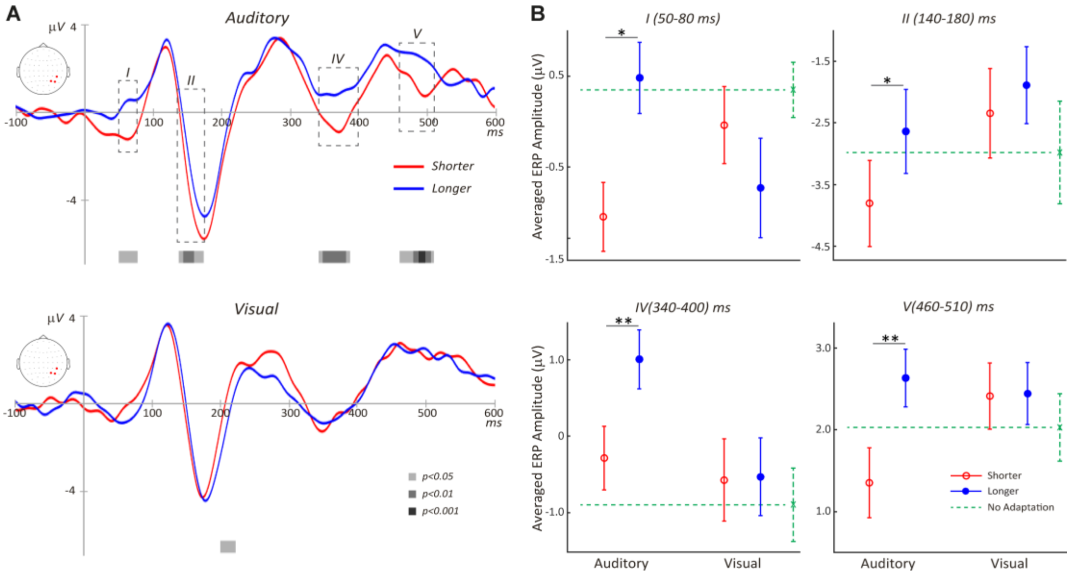


Fig. 50. Average activity across three common electrodes (P4, P6, CP6) where the auditory time interval adaptation had significant aftereffects on the ERP components. (A) Grand-averaged ERPs for different adaptation conditions. The ERPs for auditory and visual conditions are shown in separate plots (auditory: upper panel, visual: lower panel). In each plot, time-domain waveforms for shorter and longer adaptation conditions are displayed. The significant (pointwise running t-test, $p<0.05$) difference between time interval conditions are marked on the temporal axis by different gray levels. A specific time point was shaded only if at least 10 consecutive points were significantly different (see Methods). (B) Averaged ERP amplitudes in different time windows for different adaptation conditions (N=18). The temporal location of the time windows is shown in the left panel. The values for each time window are displayed in separate plots. In each plot, the open and filled symbols correspond to shorter and longer time interval conditions, respectively. Error bars indicate standard error (+ SEM) across subjects. The dotted line indicates the mean value for the no adaptation condition and the error bar placed over the symbol at the end of this line represents standard error across subjects. A significant difference between shorter and longer conditions was marked with asterisk signs (paired t-test, * $p<0.05$, **$p<0.01$).

6.3.3 Time interval aftereffects on the ERPs elicited by apparent motion: Averaged ERP amplitudes from exemplary sites

Though the significant aftereffects of auditory time interval led to different clusters for each time window, three electrodes (P4, P6, and CP6) were common across the four identified time windows. The grand-averaged ERPs of these electrodes, where the effects of auditory time interval adaptation were commonly observed, are shown in Fig. 50A. We also computed average potentials within the identified time windows for these exemplary sites and the average values are indicated by Fig. 50B. Regarding the 50-80 ms, 340-400 ms and 460-510 ms time ranges, there was a significant difference between the auditory adaptors, but not between the visual adaptors. This is confirmed by the ANOVA tests indicating a significant interaction between modality and time interval. Post-hoc pairwise comparisons indicated that the effect of time interval was only significant for auditory conditions (50-80 ms: t17 = 2.791, p = 0.012; 340-400 ms: t17 = 3.120, p = 0.006; 460-510 ms: t17 = 3.435, p = 0.003), such that the longer time interval yielded higher potentials than the shorter interval. In the 140-180 ms range, a two-way repeated measures ANOVA revealed significant main effects of modality and time interval on the ERP amplitudes, however, the interaction was not significant. The average potential for the shorter time interval was smaller than the longer interval, and the potentials for the visual stimuli were higher than for the auditory stimuli.

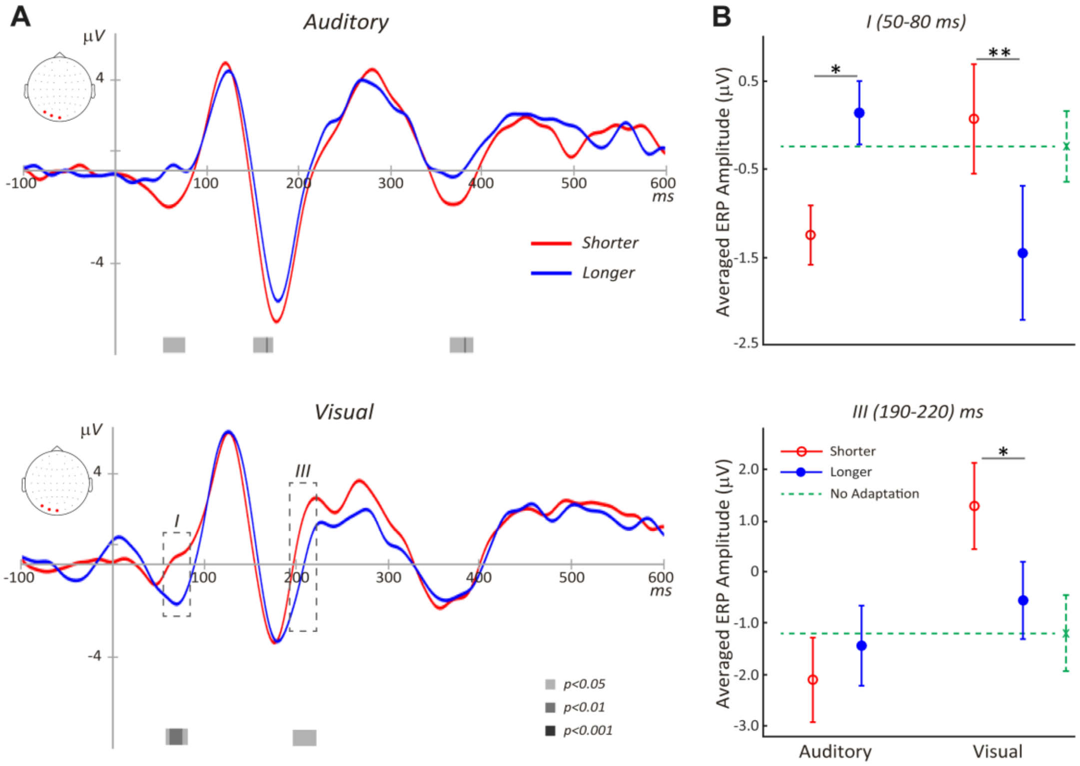


Fig. 51. Average activity across three common electrodes (O1, Oz, PO7) where the visual time interval adaptation had significant aftereffects on the ERP components. (A) Grand-averaged ERPs for different adaptation conditions. (B) Averaged ERP amplitudes in different time windows for different adaptation conditions (N=18). Other conventions are the same as those in Fig. 50.

Similarly, we identified three neighboring electrodes (O1, Oz, PO7) commonly affected by visual time intervals. The grand-averaged ERPs of these electrodes and the averaged potentials within the two time windows (50-80 ms and 190-220 ms) are shown in Fig. 51. For both time windows, a repeated measures ANOVA on the average potentials revealed a significant interaction between modality and time interval (Fig. 51B, upper panel). In the 50-80 ms range, the average potentials of the shorter were significantly higher than those of the longer time intervals for vision (t17 = -3.109, p = 0.006). On the other hand, the effect of auditory time interval was in the opposite direction (t17 = 2.702, p = 0.015). Post-hoc pairwise comparisons on the averaged potentials within 190-220 ms revealed the shorter time interval to yield higher voltages than the longer time interval for the visual adaptors (t17 = -2.585, p = 0.019), but the difference was not significant for the auditory adaptors (t17 = 1.447, p = 0.166; Fig. 51B, lower panel).

6.4 Discussion

In this experiment, we systematically examined the evoked activity to the visual motion under different time interval adaptation conditions. To define a specific time interval in the motion stimulus, we used two-frame apparent motion with a fixed time interval between the motion frames. This experiment revealed a number of interesting and novel findings. Based on the modality (auditory clicks or visual flashes) and the time interval (shorter or longer than that of the apparent motion) used during the adaptation phase, we found significant and distinct changes in ERPs elicited by exactly the same visual motion. When auditory clicks were used as the adaptors, the negative activity within 140-180 ms and the positive activity after 320 ms for shorter time interval condition were more negative and less positive than those for the longer condition mostly over parietal and parieto-central regions, respectively. On the other hand, adapting observers to visual time intervals had relatively weaker effects on the ERPs. The significant differences between shorter and longer conditions mostly clustered over occipital and parieto-occipital regions. At these scalp locations, the activity was higher for the longer condition within a shorter time range (190-320 ms). We also observed significant aftereffects of auditory and visual time intervals at common electrode locations (e.g., some of the parieto-occipital and parietal electrodes). However, at these common locations, the aftereffects occurred in distinct time windows for each modality and were in the opposite direction. Therefore, even for a common site (e.g., PO8) at which significant auditory and visual aftereffects observed, our results here indicate distinct characteristics and temporal dynamics for the aftereffects induced by each modality.

6.4.1 Sub-second Time Interval Adaptation

Our findings demonstrate for the first time that adapting to time intervals significantly alters the ERPs elicited by visual motion (i.e., a visual feature dependent on the time interval between individual frames). By indicating specific scalp sites and ERP components, our findings here suggest neural correlates underlying time interval aftereffects on motion perception. Previous psychophysical data have found significant influences of time interval adaptation on visual motion (Zhang *et al.*, 2012; Zhang & Chen, 2016). The effects of auditory and time interval adaptation were in the same direction suggesting the role for a common timing mechanism. As opposed to the view provided by these motion studies (see also Chen & Zhou, 2014), we found distinct spatiotemporal characteristics of intramodal and crossmodal time interval aftereffects. Accordingly, our results emphasize the recruitment of distinct sensory processes rather than a central supramodal network for the adaptation aftereffects on visual motion, although we cannot rule out the possibility of deep brain structures being involved in supramodal timing computations. By considering a more general argument in sub-second timing, we interpret our results as support for the distributed nature of the adaptation induced aftereffects (Murai *et al.*, 2016). Our results do not preclude a central supramodal mechanism, but in line with the general notion also suggesting the involvement of distributed processes in sub-second time interval perception.

In general, compared to vision, the auditory modality has been found to have a dominant role in temporal processing and tasks dependent on timing. Moreover, it has been shown that auditory stimuli mostly drive the perceived timing of visual events (for a review see Chen & Vroomen, 2013). Of particular interest is that learning studies revealed asymmetries between two modalities (Bratzke *et al.*, 2012; Barakat *et al.*, 2015; McGovern *et al.*, 2016). For instance, Barakat *et al.* (Barakat *et al.*, 2015) have recently reported that training observers in an auditory (but not visual) rhythm discrimination significantly improved their performance in a subsequent visual rhythm discrimination task. Similarly, using a more comprehensive design, McGovern *et al.* (2016) showed that transfer of perceptual learning between two modalities is specific to tasks and particular conditions. For temporal discrimination task, training observers in the auditory domain transferred to the visual domain (i.e., decreased discrimination thresholds in vision), but not vice versa. Even though the adaptation phase of this experiment was brief and did not include a specific learning paradigm, we also found dominant effects of auditory time interval adaptation on the evoked activity relative to visual adaptation. In this respect, our results are in line with the view supporting auditory dominance in tasks and/or stimulus attributes based on time interval and discrimination.

Contingent Negative Variation (CNV), an ERP component over frontocentral regions first described by Walter *et al.* (Walter *et al.*, 1964), has attracted great attention from researchers working on timing and time perception (van Rijn *et al.*, 2011; Kononowicz & Penney, 2016). Even though it is still subject to debate, some aspects of CNV have been found to be correlated with changes in perceptual timing (Macar *et al.*, 1999; Pfeuty *et al.*, 2008; Ng *et al.*, 2011). Strong CNV activity and correlations with timing were mostly reported by studies using a task requiring observers either to compare the time intervals (or durations) of stimuli explicitly or to reproduce the time interval (or duration) of a stimulus through a motor action. Of particular importance to this experiment, Li *et al.* (2017) have recently shown that changes in CNV can also reflect the effect of sub-second duration adaptation on the perceived duration of a subsequent event. During each trial, they first adapted observers to shorter or longer visual durations and then asked them to reproduce the duration of a visual test stimulus. Their results indicated a strong CNV component elicited by the visual test over frontocentral regions and this component was modulated by the duration used in the adaptation phase. Our results also indicated a significant negative component over frontal electrodes (Fig. 49). However, there was no significant difference between these ERPs for visual adaptation conditions. Regarding the auditory conditions, the difference was significant over only a frontal electrode (). Therefore, compared to the findings by Li *et al.* (2017), the CNV modulation was not apparent in this experiment. Here, instead of an explicit time interval discrimination or reproduction task, observers performed motion direction discrimination. Considering that the link between CNV and behavioral performance depends on cognitive context, this apparent discrepancy between the two studies may be due to the different tasks observers engaged in.

6.4.2 Auditory Timing and Visual Motion

It is well known that the temporal and spatial information provided by audition can alter visual motion perception (for reviews see Soto-Faraco *et al.*, 2003; Hidaka *et al.*, 2015). For instance, the time interval between brief stationary sounds has been shown to alter the perception of apparent motion and speed (e.g., Getzmann, 2007; Kafaligonul & Stoner, 2010, 2012). Though these behavioral studies suggest that the visual system actively uses auditory timing to estimate motion, how auditory timing is involved in shaping motion perception and the neural correlates of these influences are quite poorly understood. Unlike this line of research, in the present experiment, the auditory clicks and the temporal information were not provided concurrently with visual motion. However, any significant change in the visually evoked activity due to the auditory time interval used in the adaptation phase gives us an indication of the cortical areas actively using auditory timing. Our results indicate that the significant differences between apparent motion ERPs for shorter and longer auditory adaptation are centered over parietal electrodes. The amplitude of early negative activity (around 140-180 ms range) were significantly affected and the difference between two conditions extended over occipital scalp regions. The effects on the later (> 300 ms) positive activity were more salient and mostly centered over right parieto-central electrodes. Building from these findings, we suggest that the auditory timing may be progressively used for visual motion processing over these areas, starting from parieto-occipital up to parieto-central sites. Consistent with our observations here, visual motion areas and parietal cortices have been found to be involved in the sub-second time interval and duration percepts (Battelli *et al.*, 2007; Salvioni *et al.*, 2013; Hussain Shuler, 2016). The application of brain stimulation to right parietal cortices resulted in changes specific to sub-second duration judgments (Bueti *et al.*, 2008; Dormal *et al.*, 2016). More importantly, a neuroimaging study has shown that the activity in the right parietal cortex (i.e., Inferior Parietal Lobe) can be modulated through duration adaptation

(Hayashi *et al.*, 2015). Multisensory studies also revealed significant audiovisual interactions over parietal regions. Multisensory integration effects between two modalities were observed on the neural activity over parieto-occipital, parietal and parieto-central sites (Molholm *et al.*, 2002, 2006; Mercier *et al.*, 2013). Moreover, the amplitude negative ERP component (150-200 ms time range) over parieto-occipital sites were found to be modulated by the timing of a brief sound and these modulations were suggested to correlate with the magnitude of the temporal ventriloquism elicited by an auditory and a visual event (Stekelenburg & Vroomen, 2005). Therefore, the auditory time interval may take place for visual motion processing over these regions already known to be involved in both sub-second temporal processing and audiovisual interactions.

6.5 Limitations and Future Directions

In the current study, we used physically identical temporal intervals for both modalities. Though the difference between shorter and longer time interval conditions was perceptually obvious, they were not perceptually-equated across modalities. Since the temporal resolution of the audition is superior to vision, we expect the difference between auditory time intervals to be more salient compared to visual time intervals. Such a difference may lead to dominant influences on ERPs through auditory adaptation. On the other hand, even though we identified significant sites and time windows for visual time interval adaptation, our data and analysis may not reveal a more comprehensive and a detailed picture for visual adaptation conditions. We consider that this is possible, and characterizing the ERP component changes over a variety of timescales including supra-second time intervals is important. Determining changes for a broad range of timescales awaits further ERP investigations.

6.6 Conclusions

Taken together, the present experiment demonstrates significant aftereffects of time interval adaptation on the ERPs elicited by visual apparent motion. We found that the changes in ERPs occurred in different time windows and were mostly centered over distinct scalp sites for auditory and visual adaptors. Thus, our findings indicate a distinct change in the spatiotemporal profile of the neural activity for auditory and visual adaptation and suggest the involvement of distributed sensory processes in sub-second time interval adaptation. Moreover, the aftereffects induced by auditory stimuli were more salient compared to visual adaptors. These changes were mostly over right parietal electrodes. Accordingly, our findings here, in a combination of accumulating evidence, also support an important role for the right parietal cortices in crossmodal temporal processing.

CHAPTER 7

7 GENERAL DISCUSSIONS AND CONCLUSIONS

7.1 Discussions

Human perceptual and cognitive system has been adapted to work in a multisensory environment providing stimuli of different modalities. The multisensory processes have important roles in constructing the final perception and cognition for a successful interaction with the outside world. As mentioned in Chapter 2, although earlier studies suggest that multisensory integration is limited to high-level cortex regions and cognitive processes, recent studies proved more comprehensive view (see Fig. 52). In addition to the late interactions, many studies have recently emphasized the importance of the low-level audiovisual interactions in the final perception. For instance, it is suggested that early-level audiovisual interactions are important (or even sufficient) in explaining some of the perceptual illusions such as sound-induced flash illusion (Shams *et al.*, 2005; Mishra *et al.*, 2007, 2008). Similar findings also support the view that both early and the late interactions had a role in the final perception.

Psychophysics studies indicate that temporal asynchrony of the audiovisual stimulus is tolerated in the brain and the multisensory information is integrated is the stimuli is presented in a temporal integration window. Integration mechanisms of this integration in the brain remain unknown. We asked specific research questions about the responsible brain locations and the temporal windows of this integration. We also asked some specific research questions about the various features of the evoked activities. Our hypotheses were about the roles of the early sensory processing fields in integration and timings of the integrations of within the various locations. Our experimental designs aimed to decipher the mechanisms of integrations both in single events and also two events that create sub-second time interval or apparent motion perceptions. We also investigated the neural mechanisms of adaptation effects on apparent motion, besides the concurrent presentations of auditory and visual stimuli.

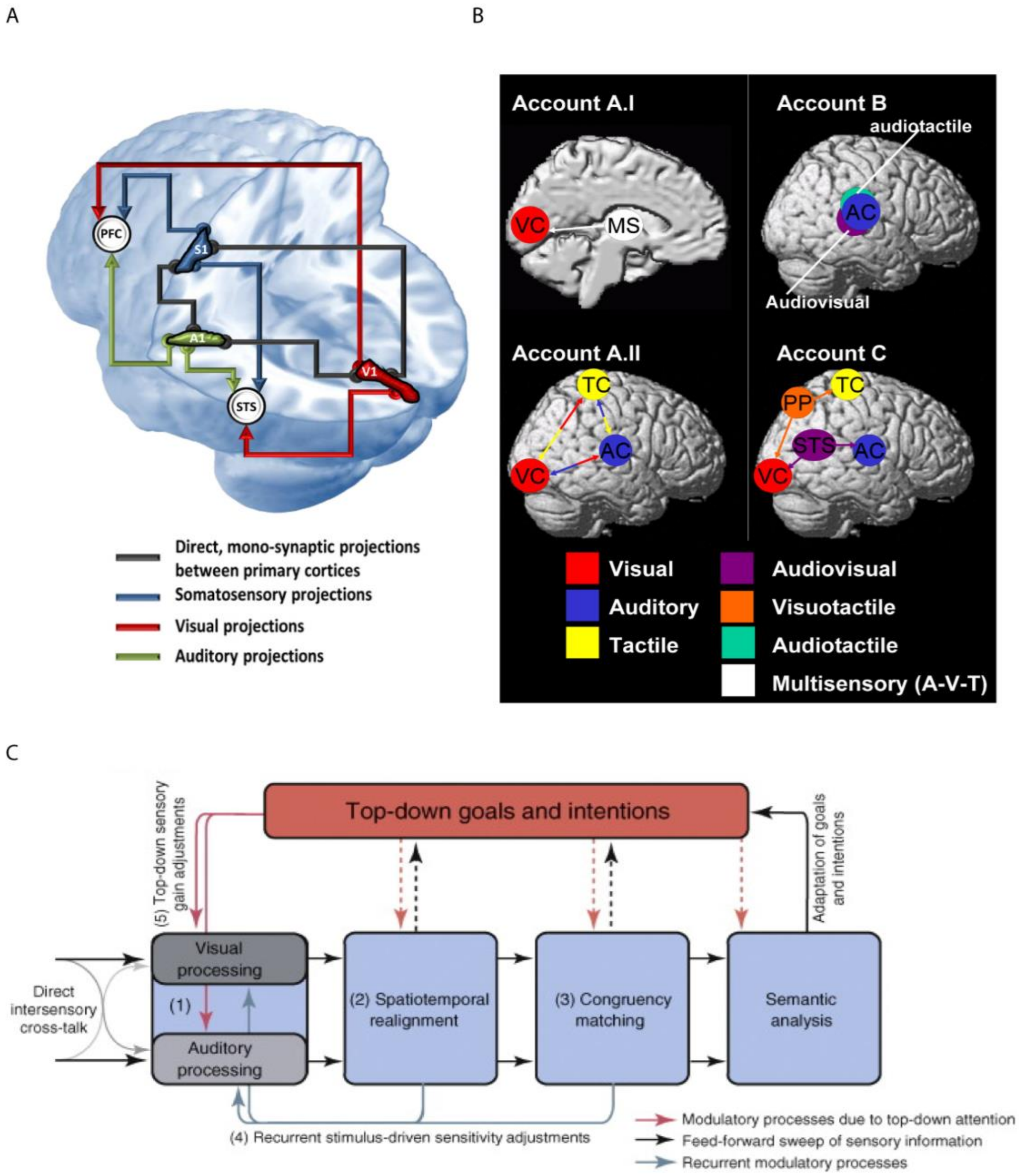


Fig. 52. (A) Demonstration of the projections having a role in the multisensory interplay between early and/or late cortical regions. V1: primary visual (red), A1: primary auditory (green), STS: superior temporal sulcus, PFC: prefrontal cortex, and their interrelationships on the left hemisphere, which play an important role in audiovisual integration (Murray *et al.*, 2016). (B) Possible neural pathways mediating multisensory interplay see Fig. 8 for other conventions. (C) A general and comprehensive framework for multisensory integration. It is suggested that both early and late sensory processes are involved in multi-sensory perception and cognition and they actively interact over the projections presented in A and B (adapted from Talsma *et al.*, 2010).

In general, late comprehensive views, as exemplified in Fig. 52, suggests that early-level processes and higher cognitive processes (such as attention, memory, and expectation) have an active role in multisensory integration. These interactive interactions are believed to occur through different cortical and subcortical areas over feedforward, recursive, and feedback connections (Fig. 52A and B). Scientific evidence indicates that to what extent these regions and connections are actively involved is altered by factors such as appropriateness between stimuli provided by different modalities, context, behavioral purpose, and task. Hence, the idea that the mechanisms of multisensory integration are used adaptively and flexibly is supported (van Atteveldt *et al.*, 2014; Talsma, 2015). For example, it has been shown that factors such as spatial distance, temporal closeness, congruence in the number of stimuli, and 'inverse effectiveness' hypothesis influence multisensory integration and manipulation in neural level as well as perceptual. Moreover, it is an open debate that cognitive loads of which the experimental paradigm requires (such as learning and attention) for different cognitive processes affect how advanced cortical areas and feedback links are involved in the integration process.

Audiovisual interactions in spatial and temporal dimensions have been studied based on the factors that potentially modulate ventriloquism such as synesthetic correspondence, attention, and other cognitive factors (Chen & Vroomen, 2013). It has been shown that the spatial ventriloquism illusion is affected both by spatial and temporal proximity between visual and auditory stimuli. An increase in spatial and temporal distance between visual and auditory stimuli has increased the likelihood that these two stimuli originate from different events, leading to a decrease in audiovisual integration (in other words, interactions) and thus a less likely perception of the illusion. On the other hand, there was no significant effect of spatial proximity on temporal ventriloquism (Vroomen & Keetels, 2006). Although the effect of temporal proximity is observed, it is known that illusions are perceived in shorter audiovisual SOAs (Morein-Zamir *et al.*, 2003). Attention mechanisms were not considered as a necessary condition for both illusions. Even in the absence of attention, spatial and temporal ventriloquism are perceived and the presence of audiovisual interactions is encountered. However, in general, it is thought that the attention mechanisms only affect and modulate these illusions (Chen & Vroomen, 2013). Therefore, slight differences between experimental designs in this thesis also aim to compare how a single event is processed when it is processed within different tasks or as being a part of different stimulus sequences. In this way, experimental designs provide a good framework to investigates the effects of the higher cognitive processes on audiovisual temporal integration.

Neural mechanisms that carry out the multisensory integration have shown in previous studies. We can group them into three accounts (Fig. 52B). We know that there are multisensory neurons as early processes as in the superior colliculus and the signals were transferred from there to the cortices (Account A.I). However, as these signals cannot be recorded by EEG, these interactions are not within the scope of this thesis. Interactions also occur between the primary sensory processing regions, as the connections in account A.II indicates. We hypothesized that we should be able to observe interactions at these regions, and the timing of the interactions would indicate the involvement of the feedforward and feedback mechanisms. The early interactions would indicate the feedforward interactions which can be accounted by this account. However, we also hypothesized the late interactions on the primary sensory regions which indicates the roles of feedback mechanisms. Additionally, effects of the multisensory regions such as PP and STS on the primary regions also considered as shown in Account C. Besides the structural organization, from the functional perspective which is presented in Fig. 52C, we also consider the top-down effects on audiovisual temporal integration. In all experiments within the scope of this thesis, since the stimulus design were congruent with each other, partial stimulus-driven processes were the same. However, our hypothesis was that both the way of the events are combined and the tasks of the experiments affect the perception of the participants. Therefore, the neural processes that are responsible for integrations of the multisensory data are also affected correspondingly. As shown in Fig. 52C top-down effects can be observed at all functional steps from early processing step of the sensory processing and integration to the late semantical steps. Besides, recurrent stimulus-driven modulatory processes also have a role in unisensory processes, to prepare and integrate the signals for the final perception that satisfy the goal of the task at semantical analysis steps.

Eventually, the aim of the present thesis was to elucidate the neural mechanisms of audiovisual temporal processing in various paradigms. Besides the vast literature in understanding the mechanisms reviewed in Chapter 2, within the scope of this thesis, our aims were: to evaluate the neural mechanisms of the auditory influences on the perception of (1) the timing of a visual stimulus, (2) the time interval defined by two consecutive visual stimuli, (3) the speed of visual apparent motion, and (4) the visual apparent motion after a time interval adaptation period. In other words, neural mechanisms of temporal ventriloquism (i.e. audiovisual interactions in the temporal dimension) were investigated. Taken together, the results from all four studies of the present thesis emphasize the robust interactions mechanisms observed in the early visual processing, while late interactions were relatively flexible and related with the temporal structure of audiovisual stimulus. Moreover, while the auditory modulation of the time interval perception is processed mostly at the early levels of the visual processing, as well as in Experiment 1; on the contrary, in apparent motion perception, auditory responses are involved also in the late visual processing steps as well as early-level visual processing. Moreover, this strong effect on the visual apparent motion was not observed only with the concurrent presentation of the auditory stimulus, but it is observed also after an auditory time interval adaptation period. Additionally, in general, low-frequency oscillatory activities recorded from various sites seem to be an essential marker for altering audiovisual temporal integration. The implications of the obtained data on the general framework mentioned above (Fig. 52) and the literature contribution are discussed below.

7.1.1 Auditory timing influences the visual processing: the role of both low-level interactions and the task

Consistent with previous studies (for review Chen & Vroomen, 2013), behavioral data shows that auditory timing has a significant influence on timing, time interval and speed perception in the concurrent presentation. Additionally, adaptation to the auditory time intervals affects the evoked potentials created by the visual apparent motion. This effect is stronger than the effect of the visual adaptation to the same stimulus. These results suggest that auditory timing information is actively used by the visual system. These results confirm the previous findings. EEG results collected in tandem have enabled us to observe these auditory timing effects in different sensory processing hierarchies. First, EEG data of all experiments show that auditory clicks evoked potentials over the early stage visual areas (Fig. 16). These potentials, which are stimulated by auditory clicks, have observed after about 80 ms after the clicks have been presented. Likewise, visual stimuli have been associated with evoked potentials on the auditory cortex (temporal region). The effects on this visual and auditory cortex indicate a fast direct interaction between low-level regions (shown in Fig. 52A, direct mono-synaptic projections between primary cortices where account A.II indicates in Fig. 52B and depicted by number 1 in the framework in Fig. 52C). Similar results were also observed in previous studies (Naue *et al.*, 2011; Thorne *et al.*, 2011; Mercier *et al.*, 2013). Such audiovisual interactions could be considered as the result of direct cortical projections between two early sensory regions, as well as through cortico-thalamic connections.

In Experiment 2 and 3, the spatial and temporal structures of the first stimulus, the first frames of the time interval and apparent motion, was the same. However, the patterns of the auditory interactions to the first presentation were different. This indicates that the difference in the task accommodates the role of the auditory stimulus differently. The evoked potential for these experiments could also be compared with evoked potentials in Experiment 1. Spatiotemporal interactions maps in all these experiments are different, although the stimulus is the same. These results support the role of the attentional mechanism on the audiovisual processing. Differences in the early time windows also indicate the role of the top-down effects on the early sensory processing steps.

7.1.2 The topography of the auditory influences on feedforward potentials alter as a binary function of SOA values, while auditory influences are distributed on feedback

This thesis aimed to evaluate the change of the scalp topography of the audiovisual temporal interactions in time with the change of SOAs between auditory and visual stimulus. The additivity analyses in Experiment 1 indicates that spatiotemporal distributions of the interactions vary according to what modality is leading. This result supports the recent findings (Cecere *et al.*, 2016, 2017). Namely, when the sound precedes the flash, there is evidence that sound prepares the occipital cortex for the forthcoming visual stimuli to process it faster and accurately (Fig. 17, Fig. 18, Fig. 19). Once interactions occur, no further interaction is observed. Which means that all the role of the auditory signal is already played and the integrated signal is transferred to the later step of the visual processes. On the other hand, when sound accompanies the visual stimulus or succeeds it, early interactions are observed in frontal regions and also further late interactions in mostly temporal regions. These findings suggest that roles are switched and visual stimulus entrains the early auditory processes and also late process over the temporal regions, as suggested before (Thorne & Debener, 2014).

In the occipital areas, two types of effect observed in time dimension (Fig. 17, Fig. 18, Fig. 19) indicates the possible roles of that early stimulus-driven recurrent sensitivity adjustments. The first interactions in 60-170-ms time windows in auditory leading conditions indicate the direct inter sensory crosstalk at the occipital region. Rhythmic change in the effect of the change in the time interval between auditory and visual stimuli shows that the inter-sensory interactions occur with a signaling mechanism that are based on the low-frequency oscillations in theta band. Secondly, the significant audiovisual interactions at a late time interval (250 ms after the visual stimulus presentation) that occurs more likely at visual leading conditions indicate recurrent effects for visual sensitivity adjustments. Another late audiovisual interaction over the frontal regions in a similar manner to the same experimental conditions indicates a distributed recurrent sensitivity adjustment over the whole brain.

Considering the general framework of work developed on multisensory integration, it is envisaged that such late interactions can be achieved through feedback and repetitive connections from the frontal and late cortical regions (for example, the ones that are shown in the Fig. 52B Account C, and depicted in Fig. 52C connections numbers 4 and 5). On the other hand, for synchrony and visual leading conditions, early frontotemporal areas showed us two types of effects. First, very early interactions within 40-100 ms and secondly (V1 to PFC and to A1 in Fig. 52A, and in Fig. 52C connections number 1), two types of late interactions in 200-250 ms range and later than 250 which are similar to the ones over posterior areas (connections number 2 and 3 in Fig. 52C). Therefore, although the auditory influences in feedforward potentials alter spatially as a binary function of SOA values, auditory influences on the feedback potentials are observed in both frontal and parietal regions, when influences are present.

We found supporting evidence also in the second and third experiments (see Fig. 29, Fig. 39, Fig. A1., Fig. A2). From this signaling point of view, distributions of the audiovisual interactions in the second and third studies can also be explained. For instance, in the longer auditory time interval conditions, first sound precedes the first flash, while second sound succeeds the second flash. Therefore, the first flash is processed faster and perceived earlier, however, sound does not precede the second flash since it was not processed faster. As a result, the interval between two flashes was perceived longer. Consequently, the level of the discrepancy between modalities automatically defines the processing speed of the first flash, hence the probability of perceiving the interval longer increases. For the shorter auditory time interval, the same rationale applies to the second flash, the one that the auditory stimulus precedes.

Such mechanism having different neural networks according to the leading sense would have an adaptive advantage from the evolutionary perspective. For auditory leading case, the kind of information transferred from auditory cortex to the early occipital region would have a predictive information about what is coming to the occipital region. When the visual stimulus fits the predictions of the auditory signals then its processing time gets faster, otherwise, feedback mechanisms play a role. This approach makes even more sense when auditory signal carries information about a visual stimulus, where ‘aboutness’ is formed by previous associations. Further research should be held to test this hypothesis that not only timing but also of the content of sound influences the speed of visual processing. Nonetheless, regardless of the content of the auditory information; in this dissertation, participants associated the click sound to the flash bar after a while, and click sound was already relevant to the visual stimulus.

7.1.3 Higher role of audition on speed perception than on time interval judgment

Recent studies highlight the auditory influences on motion perception (Hidaka *et al.*, 2015). Perceptual measurements that revealed the higher influences of auditory temporal structure on the perceived speed of the visual apparent motion compared to visual time interval judgments (Fig. 28 and Fig. 38). These differences are also observed in the auditory influences on evoked potentials. Both spatiotemporal profiles of the additivity effects and comparisons of AV-A with V responses of different conditions indicated that the spatiotemporal profile of the audiovisual interactions in the brain was more salient both in time and space in apparent motion. The auditory dominance in adaptation effects of Experiment 4 also shows that not only concurrent presentation of auditory temporal intervals but their adaptations also has a high impact on the perception of the apparent motions. This perceptual effect also observed in neural responses elicited by the apparent motion. This indicates that brain is adapted to make the auditory modality more use to determine the temporal structure of visual stimulus in the perception of the speed compared to the static visual time interval. Because temporal accuracy is probably more important for calculating the speed of the moving objects, however not for the static ones. Additionally adapting to the auditory time intervals has more effect on the apparent motion compared to visual intervals. These results support the view that mechanisms of the auditory and visual temporal processing are separated and auditory processes dominate the visual in the temporal dimension.

7.1.4 Spectral amplitude and phase are essential markers for audiovisual intersensory cross-talk

Another aim of this thesis was to investigate the roles of oscillatory mechanisms in audiovisual integration. We hypothesized that auditory influences in temporal domain can be observed in the visual-evoked spectral amplitudes, latencies, durations, and phases. In Experiment 1, spectral powers induced by unisensory and bi-sensory stimulation were observed at separate frequency bands over the occipito-parietal region. Our results in Experiment confirms the previous results (Naue *et al.*, 2011), the auditory stimulus creates gamma band activity with a stable peak latency of ~40 ms after a click onset, while visual stimulus-evoked low-beta activity with 160 ms peak latency after visual flash was presented. Furthermore, the latencies and amplitudes of low frequency evoked spectral amplitudes were changed as a function of SOA values. As previous studies investigating audiovisual integration show that auditory signals reset the phases of the ongoing signals over occipital regions and visual signals also reset the oscillations over temporal regions (Naue *et al.*, 2011; Thorne *et al.*, 2011; for a review Thorne & Debener, 2014).

The results of Experiment 1 suggested that the direct cross-talk between the auditory and visual processes are observed at the inter-trial phase coherencies of the theta band oscillations. The phase coherence analyses revealed that evoked spectral amplitudes were due to the phase resetting effects of auditory signals on the occipital regions. Therefore, auditory signal resets the ongoing low-frequency oscillations to a specific phase and oscillations evoked by the visual stimulus can either strengthen or weaken according to this phase alignment. Phase alignment is proposed to be a mechanism for interaction not only between the early sensory cortices but also is suggested as a mechanism for pre-stimulus visual awareness, originating not necessarily from the primary auditory cortex (Mathewson *et al.*, 2009, 2012; Kizuk & Mathewson, 2016). Two features of the inter-trial coherence (ITC) within 6-8 Hz frequency band changes as a function of SOA values, peak amplitudes, and latencies. First, peak amplitude changes as a sinusoidal function of SOA values. Peak to peak SOA difference was around 160 ms (see Fig. 24). Therefore, auditory stimulus resets the phase of the ~6 Hz oscillations. Auditory and visual only signals cause a phase resetting effect on the occipital channels in the similar directions and the oscillatory trend is explained by independent phase resetting effects. However, interestingly, auditory and visual peak latencies are not observed independently as two separate peaks, rather, those oscillations interacted in a way that they create a unique peak which indicates an interaction. Furthermore, the peak latencies may explain the speeded reaction at early processing steps, that no auditory interference not required. Nevertheless, 6 Hz oscillations are slow oscillations with accuracy in time is low. In conclusion, together with the ERP results, amplitude and phase analyses results provide strong evidence of auditory influences on early visual processing for the single click single flash experiments.

Oscillatory phase reset can also be observed in perceptual measurements such as saccadic reactions times (Diederich *et al.*, 2012), although reaction times were increased as a linear function of SOA values. When reaction times from individual observers are considered, oscillatory trend superimposed on the linear increase can be observed. As the previous research suggested the frequency of this oscillatory trend changes subject to subject (Diederich *et al.*, 2012).

Like the first experiment, 6-8 Hz frequency band revealed the main audiovisual interactions, also in Experiments 2 & 3. Scalp distributions of spectral amplitudes change by the length of the auditory time interval (short and long) and level of discrepancy (i.e. SOA, mild and extreme). Our results revealed that short and long auditory time interval differences observed only in the frontal region. On the other hand, mild and extreme differences were only in the posterior region. Interestingly, even though the stimuli are the same, frontal region was significantly sensitive to the conditional differences in auditory only conditions, while no significant difference observed in time interval perception experiment. This result also shows the role of attention on balancing the information processing coming from different sources. As auditory stimulus has a more influential role in speed judgment, evoked powers alter by various auditory intervals significantly in apparent motion judgment.

7.1.5 Temporal adaptation in the context of visual movement

Auditory temporal adaptation shows significant effects on apparent motion perception. In our adaptation experiment, we adapted observers to different time intervals by using either brief sounds or visual flashes and examined the evoked activity to the subsequently presented visual apparent motion. For both modalities, our results showed significant aftereffects of time interval adaptation on the ERPs elicited by visual apparent motion. However, these changes in ERPs occurred in different time windows and were mostly centered over distinct scalp sites for auditory and visual adaptors. Compared to visual adaptors, the aftereffects induced by auditory stimuli were more salient and clustered over parietal and parieto-central regions. On the other hand, adapting observers to visual time intervals had relatively weaker effects on the ERPs and mostly occurred over occipital and parieto-occipital regions. Taken together, our findings within the context of the visual motion indicate distinct changes in the spatiotemporal profile of the neural activity for auditory and visual adaptation and highlight the role of distributed sensory processes in sub-second time interval adaptation.

7.1.6 Temporal ventriloquism and cognitive processes

As mentioned in previous chapters, many studies in the literature also point to the effects of advanced cognitive processes on multisensory integration and motion perception (Talsma, 2015). For example, it has been observed that there is a significant impact on visual motion perception and sensitivity to the motion after associating stationary tones with no motion information (Hidaka *et al.*, 2011; Kafaligonul & Oluk, 2015). These results highlight not only the role of physical features provided by auditory stimuli but also the role of auditory stimuli in the perception of the apparent motion when associated with visual motion. Differences in the evoked responses to the first visual frames indicate the task depended on pre-stimulus activities and more importantly the mechanisms of the allocation of the auditory temporal information in visual processing. In addition, contrary to the general belief, some studies (e.g., Freeman & Driver, 2008; Donohue *et al.*, 2015) suggest that attention is not required to achieve audiovisual interactions in time, it is emphasized that attention may have a modulating effect on temporal ventriloquism (Chen & Vroomen, 2013). Like many similar studies (Naue *et al.*, 2011; Thorne *et al.*, 2011), in the experiments here, participants were instructed to ignore these stimuli if any auditory stimuli were presented while performing tasks based on visual stimulation. A study with controlled attentional mechanism will be very helpful in direct research on the role of attention in the audiovisual temporal processing and speed perception.

7.2 Concluding Remarks and Future Directions

Finally, this thesis aimed to investigate the neural mechanisms of the sub-second audiovisual temporal integration within the scope of the single event, time interval and apparent motion perceptions. Additionally, effects of the auditory and visual time interval adaptation on the evoked potentials elicited by visual apparent motion are investigated. Within the scope of this dissertation, we can affirm that (1) as inferred from the Chapters 3, 4 & 5, spatiotemporal profile of audiovisual interactions are not only affected by the stimulus properties, but the tasks also influence how the temporal structure of the auditory stimulus is allocated in visual processing; (2) Chapter 3 shows that depending on the priority at the presentation time, early stimulus of one modality plays a precursor role for the other sensory processing network, specific to the task of the experiment, auditory stimulus prepared the visual region by the phase resetting mechanisms at theta band; (3) as Chapter 4 & 5 show that temporal structure of the auditory modality has more influence on the visual apparent motion compared to time interval perception, and this auditory influence on visual perception is observed both in ERPs and theta band spectral powers which are distributed over time and location; (4) as Chapter 5 & 6 shows temporal structure of the auditory stimulus strongly influences the visual apparent motion, and the interactions distributed in time with the important role of right parietal region; (5) As spectral analyses in Chapter 3,4 & 5 indicates, multisensory cross-talks occurs at theta band, specifically over the phase components locked to the stimulus presentation.

As a next step, for the sake of the generalizability of the robust results about the neural mechanisms, we will consider using naturalistic stimuli with the similar temporal structures. As the Experiment 1 suggests the auditory stimulus a precursor role, content of the auditory stimulus should also have an influence on the speed of the processing the visual stimulus. Therefore, the effect of the ‘aboutness’ of the auditory stimulus will be investigated in behavioral level, and if present, neural mechanisms will of the differences will also be investigated further. Analyses of the Experiment 2 & 3, will also be improved to identify the roles of early interactions of single audiovisual integration and the late influences of the auditory time interval on the time interval and apparent motion. And analyses will also be improved in order to leave the possible confounding effects out due to the comparisons of two sequentially presented time intervals. Additionally, experimental paradigms used in this dissertation that shows robust behavioral effects will also be tested with the bipolar and schizophrenia patients whose multisensory integration mechanisms operate differently. These tests will be used to investigate the neural mechanisms of the disorder.

REFERENCES

Alais, D. & Burr, D. (2004a) The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.*, **14**, 257–262.

Alais, D. & Burr, D. (2004b) No direction-specific bimodal facilitation for audiovisual motion detection. *Cogn. Brain Res.*, **19**, 185–194.

Allen, P.J., Polizzi, G., Krakow, K., Fish, D.R., & Lemieux, L. (1998) Identification of EEG Events in the MR Scanner: The Problem of Pulse Artifact and a Method for Its Subtraction. *Neuroimage*, **8**, 229–239.

Avillac, M., Ben Hamed, S., & Duhamel, J.-R. (2007) Multisensory Integration in the Ventral Intraparietal Area of the Macaque Monkey. *J. Neurosci.*, **27**.

Barakat, B., Seitz, A.R., & Shams, L. (2015) *Visual Rhythm Perception Improves through Auditory but Not Visual Training*, Current Biology. Elsevier.

Barraclough*, N.E., Xiao*, D., Baker, C.I., Oram, M.W., & Perrett, D.I. (2005) Integration of Visual and Auditory Information by Superior Temporal Sulcus Neurons Responsive to the Sight of Actions. *J. Cogn. Neurosci.*, **17**, 377–391.

Barth, D.S., Goldberg, N., Brett, B., & Di, S. (1995) The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Res.*, **678**, 177–190.

Başar-Eroglu, C., Strüber, D., Schürmann, M., Stadler, M., & Başar, E. (1996) Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance. *Int. J. Psychophysiol.*, **24**, 101–112.

Battelli, L., Pascual-Leone, A., & Cavanagh, P. (2007) The “when” pathway of the right parietal lobe. *Trends Cogn. Sci.*, **11**, 204–210.

Beauchamp, M.S., Lee, K.E., Argall, B.D., & Martin, A. (2004) Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. *Neuron*, **41**, 809–823.

Benevento, L.A., Fallon, J., Davis, B.J., & Rezak, M. (1977) Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Exp. Neurol.*, **57**, 849–872.

Bental, E., Dafny, N., & Feldman, S. (1968) Convergence of auditory and visual stimuli on single cells in the primary visual cortex of unanesthetized unrestrained cats. *Exp. Neurol.*, **20**, 341–351.

Bertelson, P. & Aschersleben, G. (1998) Automatic visual bias of perceived auditory location. *Psychon. Bull. Rev.*, **5**, 482–489.

Brainard, D.H. (1997) The psychophysics toolbox. *Spat. Vis.*, **10**, 433–436.

Bratzke, D., Seifried, T., & Ulrich, R. (2012) Perceptual learning in temporal discrimination: asymmetric cross-modal transfer from audition to vision. *Exp.*

Brain Res., **221**, 205–210.

Bruce, C., Desimone, R., & Gross, C.G. (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.*, **46**.

Bruno, A. & Cicchini, G.M. (2016) Multiple channels of visual time perception. *Curr. Opin. Behav. Sci.*, **8**, 131–139.

Bueti, D., Bahrami, B., & Walsh, V. (2008) Sensory and Association Cortex in Time Perception. *J. Cogn. Neurosci.*, **20**, 1054–1062.

Burr, D., Banks, M.S., & Morrone, M.C. (2009) Auditory dominance over vision in the perception of interval duration. *Exp. brain Res.*, **198**, 49–57.

Busch, C., Wilson, G., Orr, C., & Papanicolaou, A. (1989) Crossmodal Interactions of Auditory Stimulus Presentation on the Visual Evoked Magnetic Response. In Williamson, S.J., Hoke, M., Stroink, G., & Kotani, M. (eds), *Advances in Biomagnetism*. Springer US, Boston, MA, pp. 221–224.

Busse, L., Roberts, K.C., Crist, R.E., Weissman, D.H., & Woldorff, M.G. (2005) The spread of attention across modalities and space in a multisensory object. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 18751–18756.

Buzsáki, G. & Buzsaki, G. (2006) *Rhythms of the Brain*, Rhythms of the Brain. Oxford University Press.

Callaway, E. & Yeager, C.L. (1960) Relationship between reaction time and electroencephalographic alpha phase. *Science*, **132**, 1765–1766.

Calvert, G.A., Campbell, R., & Brammer, M.J. (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.*, **10**, 649–657.

Calvert, G., Spence, C., & Stein, B. (2004) *The Handbook of Multisensory Processes*.

Cecere, R., Gross, J., & Thut, G. (2016) Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *Eur. J. Neurosci.*, **43**, 1561–1568.

Cecere, R., Gross, J., Willis, A., & Thut, G. (2017) Being First Matters: Topographical Representational Similarity Analysis of ERP Signals Reveals Separate Networks for Audiovisual Temporal Binding Depending on the Leading Sense. *J. Neurosci.*, **37**, 5274–5287.

Chen, L. & Vroomen, J. (2013) Intersensory binding across space and time: a tutorial review. *Atten. Percept. Psychophys.*, **75**, 790–811.

Chen, L. & Zhou, X. (2014) Fast transfer of crossmodal time interval training. *Exp. Brain Res.*, **232**, 1855–1864.

Cohen, M.X. (2017) Where Does EEG Come From and What Does It Mean? *Trends Neurosci.*, **40**, 208–218.

Diederich, A., Schomburg, A., & Colonius, H. (2012) Saccadic Reaction Times to Audiovisual Stimuli Show Effects of Oscillatory Phase Reset. *PLoS One*, **7**, e44910.

Ding, Y., Martinez, A., Qu, Z., & Hillyard, S.A. (2014) Earliest stages of visual cortical processing are not modified by attentional load. *Hum. Brain Mapp.*, **35**, 3008–3024.

Donohue, S.E., Green, J.J., & Woldorff, M.G. (2015) The effects of attention on the temporal integration of multisensory stimuli. *Front. Integr. Neurosci.*, **9**, 32.

Dormal, V., Javadi, A.-H., Pesenti, M., Walsh, V., & Cappelletti, M. (2016) Enhancing duration processing with parietal brain stimulation. *Neuropsychologia*, **85**, 272–277.

Driver, J. & Noesselt, T. (2008) Multisensory Interplay Reveals Crossmodal Influences on "Sensory-Specific" Brain Regions, Neural Responses, and Judgments. *Neuron*, **57**, 11–23.

Driver, J. & Spence, C. (1998) Crossmodal attention. *Curr. Opin. Neurobiol.*, **8**, 245–253.

Dustman, R.E. & Beck, E.C. (1965) Phase of alpha brain waves, reaction time and visually evoked potentials. *Electroencephalogr. Clin. Neurophysiol.*, **18**, 433–440.

Escoffier, N., Herrmann, C.S., & Schirmer, A. (2015) Auditory rhythms entrain visual processes in the human brain: evidence from evoked oscillations and event-related potentials. *Neuroimage*, **111**, 267–276.

Felleman, D.J. & Van Essen, D.C. (1991) Distributed Hierarchical Processing in the Cerebral Cortex. *Cereb. Cortex*, **1**, 1–47.

Fendrich, R. & Corballis, P.M. (2001) The temporal cross-capture of audition and vision. *Percept. Psychophys.*, **63**, 719–725.

Feng, W., Störmer, V.S., Martinez, A., McDonald, J.J., & Hillyard, S.A. (2014) Sounds activate visual cortex and improve visual discrimination. *J. Neurosci.*, **34**, 9817–9824.

Fiebelkorn, I.C., Foxe, J.J., Butler, J.S., Mercier, M.R., Snyder, A.C., & Molholm, S. (2011) Ready, Set, Reset: Stimulus-Locked Periodicity in Behavioral Performance Demonstrates the Consequences of Cross-Sensory Phase Reset. *J. Neurosci.*, **31**.

Fiebelkorn, I.C.C., Snyder, A.C.C., Mercier, M.R.R., Butler, J.S.S., Molholm, S., & Foxe, J.J.J. (2013) Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage*, **69**, 126–137.

Fisher, N.I. (1995) *Statistical Analysis of Circular Data*. Cambridge University Press.

Fishman, M.C. & Michael, P. (1973) Integration of auditory information in the cat's visual cortex. *Vision Res.*, **13**, 1415–1419.

Foxe, J.J. & Schroeder, C.E. (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, **16**, 419–423.

Freeman, E. & Driver, J. (2008) Direction of visual apparent motion driven solely by timing of a static sound. *Curr. Biol.*, **18**, 1262–1266.

Frens, M.A. & Van Opstal, A.J. (1995) A quantitative study of auditory-evoked saccadic eye movements in two dimensions. *Exp. Brain Res.*, **107**, 103–117.

Fründ, I., Schadow, J., Busch, N.A., Körner, U., & Herrmann, C.S. (2007) Evoked γ oscillations in human scalp EEG are test–retest reliable. *Clin. Neurophysiol.*, **118**, 221–227.

Fuster, J.M., Bodner, M., & Kroger, J.K. (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, **405**, 347–351.

Gebhard, J.W. & Mowbray, G.H. (1959) On discriminating the rate of visual flicker and auditory flutter. *Am. J. Psychol.*, **72**, 521–529.

Getzmann, S. (2007) The Effect of Brief Auditory Stimuli on Visual Apparent Motion. *Perception*, **36**, 1089–1103.

Ghazanfar, A.A. & Schroeder, C.E. (2006) Is neocortex essentially multisensory? *Trends Cogn. Sci.*, **10**, 278–285.

Gielen, S.C.A.M., Schmidt, R.A., & Van Den Heuvel, P.J.M. (1983) On the nature of intersensory facilitation of reaction time. *Percept. Psychophys.*, **34**, 161–168.

Grahn, J.A., Henry, M.J., & McAuley, J.D. (2011) FMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. *Neuroimage*, **54**, 1231–1243.

Graziano, M., Yap, G., & Gross, C. (1994) Coding of visual space by premotor neurons. *Science (80-.).*, **266**.

Graziano, M.S.A., Reiss, L.A.J., & Gross, C.G. (1999) A neuronal representation of the location of nearby sounds. *Nature*, **397**, 428–430.

Grondin, S. (2010) Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, Psychophys.*, **72**, 561–582.

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2002) Audiotactile interactions in roughness perception. *Exp. Brain Res.*, **146**, 161–171.

Guthrie, D. & Buchwald, J.S. (1991) Significance testing of difference potentials. *Psychophysiology*, **28**, 240–244.

Hayashi, M.J., Ditye, T., Harada, T., Hashiguchi, M., Sadato, N., Carlson, S., Walsh, V., & Kanai, R. (2015) Time Adaptation Shows Duration Selectivity in the Human Parietal Cortex. *PLOS Biol.*, **13**, e1002262.

Herculano-Houzel, S. (2009) The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.*, **3**, 31.

Hershenson, M. (1962) Reaction time as a measure of intersensory facilitation. *J. Exp. Psychol.*, **63**, 289–293.

Hidaka, S., Teramoto, W., Kobayashi, M., & Sugita, Y. (2011) Sound-contingent visual motion aftereffect. *BMC Neurosci.*, **12**, 44.

Hidaka, S., Teramoto, W., & Sugita, Y. (2015) Spatiotemporal Processing in Crossmodal Interactions for Perception of the External World: A Review. *Front. Integr. Neurosci.*, **9**, 62.

Hikosaka, K., Iwai, E., Saito, H., & Tanaka, K. (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *J. Neurophysiol.*, **60**, 1615–1637.

Howard, I.P. & Templeton, W.B. (1966) *Human Spatial Orientation*, Human spatial orientation. Wiley, New York.

Hussain Shuler, M.G. (2016) Timing in the visual cortex and its investigation. *Curr. Opin. Behav. Sci.*, **8**, 73–77.

Jones, E.G. & Powell, T.P.S. (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, **93**, 793–820.

Jousmäki, V., Hari, R., & Stein, B. (1998) Parchment-skin illusion: sound-biased touch. *Curr. Biol.*, **8**, R190.

Kafaligonul, H. & Oluk, C. (2015) Audiovisual associations alter the perception of low-level visual motion. *Front. Integr. Neurosci.*, **9**, 26.

Kafaligonul, H. & Stoner, G.R. (2010) Auditory modulation of visual apparent motion with short spatial and temporal intervals. *J. Vis.*, **10**, 31.

Kafaligonul, H. & Stoner, G.R. (2012) Static sound timing alters sensitivity to low-level visual motion. *J. Vis.*, **12**, 2-.

Kanai, R., Lloyd, H., Bueti, D., & Walsh, V. (2011) Modality-independent role of the primary auditory cortex in time estimation. *Exp. brain Res.*, **209**, 465–471.

Kaya, U., Yildirim, F.Z., & Kafaligonul, H. (2017) The involvement of centralized and distributed processes in sub-second time interval adaptation: an ERP investigation of apparent motion. *Eur. J. Neurosci.*, **46**, 2325–2338.

Kayser, C. & Logothetis, N.K. (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.*, **212**, 121–132.

Kayser, C., Petkov, C.I., & Logothetis, N.K. (2008) Visual modulation of neurons in auditory cortex. *Cereb. Cortex*, **18**, 1560–1574.

Keil, J., Pomper, U., Feuerbach, N., & Senkowski, D. (2017) Temporal orienting precedes intersensory attention and has opposing effects on early evoked brain activity. *Neuroimage*,.

Kizuk, S.A.D. & Mathewson, K.E. (2016) Power and Phase of Alpha Oscillations Reveal an Interaction between Spatial and Temporal Visual Attention. *J. Cogn. Neurosci.*, 1–15.

Kononowicz, T.W. & Penney, T.B. (2016) The contingent negative variation (CNV): timing isn't everything. *Curr. Opin. Behav. Sci.*, **8**, 231–237.

Lakatos, P., Chen, C.-M., O'Connell, M.N., Mills, A., & Schroeder, C.E. (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, **53**, 279–292.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., & Schroeder, C.E. (2008) Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science (80-.).*, **320**.

Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., & Schroeder, C.E. (2013) The Spectrotemporal Filter Mechanism of Auditory Selective Attention. *Neuron*, **77**, 750–761.

Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., & Schroeder, C.E. (2009) The Leading Sense: Supramodal Control of Neurophysiological Context by Attention. *Neuron*, **64**, 419–430.

Levitan, C.A., Ban, Y.-H.A., Stiles, N.R.B., & Shimojo, S. (2015) Rate perception adapts across the senses: evidence for a unified timing mechanism. *Sci. Rep.*, **5**, 8857.

Li, B., Chen, Y., Xiao, L., Liu, P., & Huang, X. (2017) Duration adaptation modulates EEG correlates of subsequent temporal encoding. *Neuroimage*, **147**, 143–151.

Li, B., Yuan, X., & Huang, X. (2015) The aftereffect of perceived duration is contingent on auditory frequency but not visual orientation. *Sci. Rep.*, **5**, 10124.

Linares, D., Cos, I., & Roseboom, W. (2016) Adaptation for multisensory relative timing. *Curr. Opin. Behav. Sci.*, **8**, 35–41.

Macar, F., Vidal, F., & Casini, L. (1999) The supplementary motor area in motor and sensory timing: evidence from slow brain potential changes. *Exp. brain Res.*, **125**, 271–280.

Mathewson, K.E., Gratton, G., Fabiani, M., Beck, D.M., & Ro, T. (2009) To See or Not to See: Prestimulus Phase Predicts Visual Awareness. *J. Neurosci.*, **29**, 2725–2732.

Mathewson, K.E., Prudhomme, C., Fabiani, M., Beck, D.M., Lleras, A., & Gratton, G. (2012) Making waves in the stream of consciousness: entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. *J. Cogn. Neurosci.*, **24**, 2321–2333.

Mauk, M.D. & Buonomano, D. V (2004) The neural basis of temporal processing. *Annu. Rev. Neurosci.*, **27**, 307–340.

Mazaheri, A., van Schouwenburg, M.R., Dimitrijevic, A., Denys, D., Cools, R., & Jensen, O. (2013) Region specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. *Neuroimage*, **87**, 356–362.

McDonald, J.J., Störmer, V.S., Martinez, A., Feng, W., & Hillyard, S.A. (2013) Salient sounds activate human visual cortex automatically. *J. Neurosci.*, **33**, 9194–9201.

McDonald, J.J., Teder-Sälejärvi, W.A., & Hillyard, S.A. (2000) Involuntary orienting to sound improves visual perception. *Nature*, **407**, 906–908.

McGovern, D.P., Astle, A.T., Clavin, S.L., & Newell, F.N. (2016) Task-specific transfer of perceptual learning across sensory modalities. *Curr. Biol.*, **26**, R20–R21.

McGurk, H. & Macdonald, J. (1976) Hearing lips and seeing voices. *Nature*, **264**, 691–811.

Meck, W.H. & Ivry, R.B. (2016) Editorial overview: Time in perception and action. *Curr. Opin. Behav. Sci.*, **8**, vi–x.

Merabet, L.B. & Pascual-Leone, A. (2010) Neural reorganization following sensory loss: the opportunity of change. *Nat. Rev. Neurosci.*, **11**, 44–52.

Merchant, H. & de Lafuente, V. (2014) Introduction to the Neurobiology of Interval Timing. Springer New York, pp. 1–13.

Merchant, H., Harrington, D.L., & Meck, W.H. (2013) Neural Basis of the Perception and Estimation of Time. *Annu. Rev. Neurosci.*, **36**, 313–336.

Mercier, M.R., Foxe, J.J., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H., & Molholm, S. (2013) Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. *Neuroimage*, **79**, 19–29.

Meredith, M.A. (2002) On the neuronal basis for multisensory convergence: a brief overview. *Cogn. Brain Res.*, **14**, 31–40.

Meredith, M.A., Allman, B.L., Keniston, L.P., & Clemo, H.R. (2012) Are Bimodal Neurons the Same throughout the Brain? In Murray MM & Wallace MT (eds), *The Neural Bases of Multisensory Processes*. CRC Press/Taylor & Francis, Boca Raton (FL).

Miller, R.L., Stein, B.E., & Rowland, B.A. (2017) Multisensory Integration Uses a Real-Time Unisensory–Multisensory Transform. *J. Neurosci.*, **37**.

Mishra, J., Martinez, A., & Hillyard, S.A. (2008) Cortical processes underlying sound-induced flash fusion. *Brain Res.*, **1242**, 102–115.

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. *J. Neurosci.*, **27**, 4120–4131.

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. *Cogn. Brain Res.*, **14**, 115–128.

Molholm, S., Sehatpour, P., Mehta, A.D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J.P., Schwartz, T.H., & Foxe, J.J. (2006) Audio-Visual Multisensory Integration in Superior Parietal Lobule Revealed by Human Intracranial Recordings. *J. Neurophysiol.*, **96**, 721–729.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003) Auditory capture of vision: Examining temporal ventriloquism. *Cogn. Brain Res.*, **17**, 154–163.

Morrell, F. (1972) Visual system's view of acoustic space. *Nature*, **238**, 44–46.

Murai, Y., Whitaker, D., & Yotsumoto, Y. (2016) The centralized and distributed nature of adaptation-induced misjudgments of time. *Curr. Opin. Behav. Sci.*, **8**, 117–123.

Murata, K., Cramer, H., & Bach-y-Rita, P. (1965) Neuronal convergence of noxious, acoustic, and visual stimuli in the visual cortex of the cat. *J. Neurophysiol.*, **28**, 1223–1239.

Murray, M.M., Lewkowicz, D.J., Amedi, A., & Wallace, M.T. (2016) Multisensory Processes: A Balancing Act across the Lifespan. *Trends Neurosci.*, **39**, 567–579.

Naue, N., Rach, S., Strüber, D., Huster, R.J., Zaehle, T., Körner, U., & Herrmann, C.S. (2011) Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *J. Neurosci.*, **31**, 7729–7736.

Ng, K.K., Tobin, S., & Penney, T.B. (2011) Temporal Accumulation and Decision Processes in the Duration Bisection Task Revealed by Contingent Negative Variation. *Front. Integr. Neurosci.*, **5**, 77.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.*, **2011**, 156869.

Pelli, D.G. (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.*, **10**, 437–442.

Perrault, T.J., Vaughan, J.W., Stein, B.E., & Wallace, M.T. (2003) Neuron-Specific Response Characteristics Predict the Magnitude of Multisensory Integration. *J. Neurophysiol.*, **90**.

Petro, L.S., Paton, A.T., & Muckli, L. (2017) Contextual modulation of primary visual cortex by auditory signals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **372**.

Pfeuty, M., Ragot, R., & Pouthas, V. (2008) Brain activity during interval timing depends on sensory structure. *Brain Res.*, **1204**, 112–117.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., & Taylor, M.J. (2000) Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, **37**, 127–152.

Rammsayer, T.H., Borter, N., & Troche, S.J. (2015) Visual-auditory differences in duration discrimination of intervals in the subsecond and second range. *Front. Psychol.*, **6**, 1626.

Recanzone, G.H. (2003) Auditory influences on visual temporal rate perception. *J. Neurophysiol.*, **89**, 1078–1093.

Recanzone, G.H. (2009) Interactions of auditory and visual stimuli in space and time. *Hear. Res.*, **258**, 89–99.

Romei, V., Gross, J., & Thut, G. (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. *Curr. Biol.*, **22**, 807–813.

Saito, D.N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., & Sadato, N. (2005) Cross-modal Binding and Activated Attentional Networks during Audio-visual Speech Integration: a Functional MRI Study. *Cereb. Cortex*, **15**, 1750–1760.

Salvioni, P., Murray, M.M., Kalmbach, L., & Bueti, D. (2013) How the Visual Brain Encodes and Keeps Track of Time. *J. Neurosci.*, **33**, 12423–12429.

Scheier, C., Nijhawan, R., & Shimojo, S. (1999) Sound alters visual temporal resolution. *Invest. Ophthalmol. Vis. Sci.*, **40**.

Schroeder, C., Molhom, S., Lakatos, P., Ritter, W., & Foxe, J. (2004) Human-simian correspondence in the early cortical processing of multisensory cues. *Cogn. Process.*, **5**, 140–151.

Sekuler, R., Sekuler, A.B., & Lau, R. (1997) Sound alters visual motion perception. *Nature*, **385**, 308.

Senkowski, D., Molholm, S., Gomez-Ramirez, M., & Foxe, J.J. (2005) Oscillatory Beta Activity Predicts Response Speed during a Multisensory Audiovisual Reaction Time Task: A High-Density Electrical Mapping Study. *Cereb. Cortex*, **16**, 1556–1565.

Senkowski, D., Saint-Amour, D., Kelly, S.P., & Foxe, J.J. (2007) Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study. *Neuroimage*, **36**, 877–888.

Senkowski, D., Schneider, T.R., Foxe, J.J., & Engel, A.K. (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci.*, **31**, 401–409.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005) Early modulation of visual cortex by sound: an MEG study. *Neurosci. Lett.*, **378**, 76–81.

Shams, L., Kamitani, Y., & Shimojo, S. (2000) Illusions. What you see is what you hear. *Nature*, **408**, 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002) Visual illusion induced by sound. *Cogn. Brain Res.*, **14**, 147–152.

Shi, Z., Chen, L., & Müller, H.J. (2010) Auditory temporal modulation of the visual Ternus effect: the influence of time interval. *Exp. brain Res.*, **203**, 723–735.

Shimojo, S. & Shams, L. (2001) Sensory modalities are not separate modalities:

plasticity and interactions. *Curr. Opin. Neurobiol.*, **11**, 505–509.

Soto-Faraco, S., Kingstone, A., & Spence, C. (2003) Multisensory contributions to the perception of motion. *Neuropsychologia,* **41**, 1847–1862.

Spinelli, D.N., Starr, A., & Barrett, T.W. (1968) Auditory specificity in unit recordings from cat's visual cortex. *Exp. Neurol.*, **22**, 75–84.

Staal, H.E. & Donderi, D.C. (1983) The effect of sound on visual apparent movement. *Am. J. Psychol.*, **96**, 95–105.

Stein, B.E. (1998) Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Exp. brain Res.*, **123**, 124–135.

Stein, B.E. & Arigbede, M.O. (1972) Unimodal and multimodal response properties of neurons in the cat's superior colliculus. *Exp. Neurol.*, **36**, 179–196.

Stein, B.E. & Meredith, M.E. (1993) *The Merging of the Sense*. The MIT Press.

Stein, B.E., Stanford, T.R., Ramachandran, R., Perrault, T.J., & Rowland, B.A. (2009) Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Exp. Brain Res.*, **198**, 113–126.

Stein, B.E. & Wallace, M.T. (1996) Comparisons of cross-modality integration in midbrain and cortex. *Prog. Brain Res.*, **112**, 289–299.

Stekelenburg, J.J. & Vroomen, J. (2005) An event-related potential investigation of the time-course of temporal ventriloquism. *Neuroreport,* **16**, 641–644.

Stekelenburg, J.J. & Vroomen, J. (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.*, **19**, 1964–1973.

Stekelenburg, J.J. & Vroomen, J. (2012) Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Front. Integr. Neurosci.*, **6**, 26.

Sugihara, T., Diltz, M.D., Averbeck, B.B., & Romanski, L.M. (2006) Integration of Auditory and Visual Communication Information in the Primate Ventrolateral Prefrontal Cortex. *J. Neurosci.*, **26**.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996) Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J. Neurosci.*, **16**, 4240–4249.

Talsma, D. (2015) Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Front. Integr. Neurosci.*, **9**.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M.G. (2010) The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.*, **14**, 400–410.

Thorne, J.D., De Vos, M., Viola, F.C., & Debener, S. (2011) Cross-modal phase reset predicts auditory task performance in humans. *J. Neurosci.*, **31**, 3853–3861.

Thorne, J.D. & Debener, S. (2014) Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.*, **307**, 144–152.

van Atteveldt, N., Murray, M.M.M., Thut, G., Schroeder, C.E.E., van Atteveldt, N., Murray, M.M.M., Thut, G., Schroeder, C.E.E., Atteveldt, N. van, Murray, M.M.M., Thut, G., Schroeder, C.E.E., van Atteveldt, N., Murray, M.M.M., Thut, G., & Schroeder, C.E.E. (2014) Multisensory integration: flexible use of general operations. *Neuron*, **81**, 1240–1253.

van Rijn, H., Kononowicz, T.W., Meck, W.H., Ng, K.K., & Penney, T.B. (2011) Contingent negative variation and its relation to time estimation: a theoretical evaluation. *Front. Integr. Neurosci.*, **5**, 91.

Vaudano, E., Legg, C.R., & Glickstein, M. (1991) Afferent and Efferent Connections of Temporal Association Cortex in the Rat: A Horseradish Peroxidase Study. *Eur. J. Neurosci.*, **3**, 317–330.

Vidal, M., Whitaker, D., McGraw, P., Suzuki, S., & Boon, R. (2017) Hearing flashes and seeing beeps: Timing audiovisual events. *PLoS One*, **12**, e0172028.

Vroomen, J. & de Gelder, B. (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.*, **26**, 1583–1590.

Vroomen, J. & Keetels, M. (2006) The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *J. Exp. Psychol. Hum. Percept. Perform.*, **32**, 1063–1071.

Vroomen, J. & Stekelenburg, J.J. (2010) Visual Anticipatory Information Modulates Multisensory Interactions of Artificial Audiovisual Stimuli. *J. Cogn. Neurosci.*, **22**, 1583–1596.

Walter, W.G., Cooper, R., Aldridge, V.J., McCallum, W.C., & Winter, A.L. (1964) Contingent Negative Variation : An Electric Sign of Sensori-Motor Association and Expectancy in the Human Brain. *Nature*, **203**, 380–384.

Welch, R. & Warren, D. (1980) Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.*,.

Welch, R.B., DutionHurt, L.D., & Warren, D.H. (1986) Contributions of audition and vision to temporal rate perception. *Percept. Psychophys.*, **39**, 294–300.

Zhang, H., Chen, L., & Zhou, X. (2012) Adaptation to visual or auditory time intervals modulates the perception of visual apparent motion. *Front. Integr. Neurosci.*, **6**, 100.

Zhang, Y. & Chen, L. (2016) Crossmodal Statistical Binding of Temporal Information and Stimuli Properties Recalibrates Perception of Visual Apparent Motion. *Front. Psychol.*, **7**, 434.

APPENDIX

Additivity analyses of all channels and conditions for Experiment 2&3

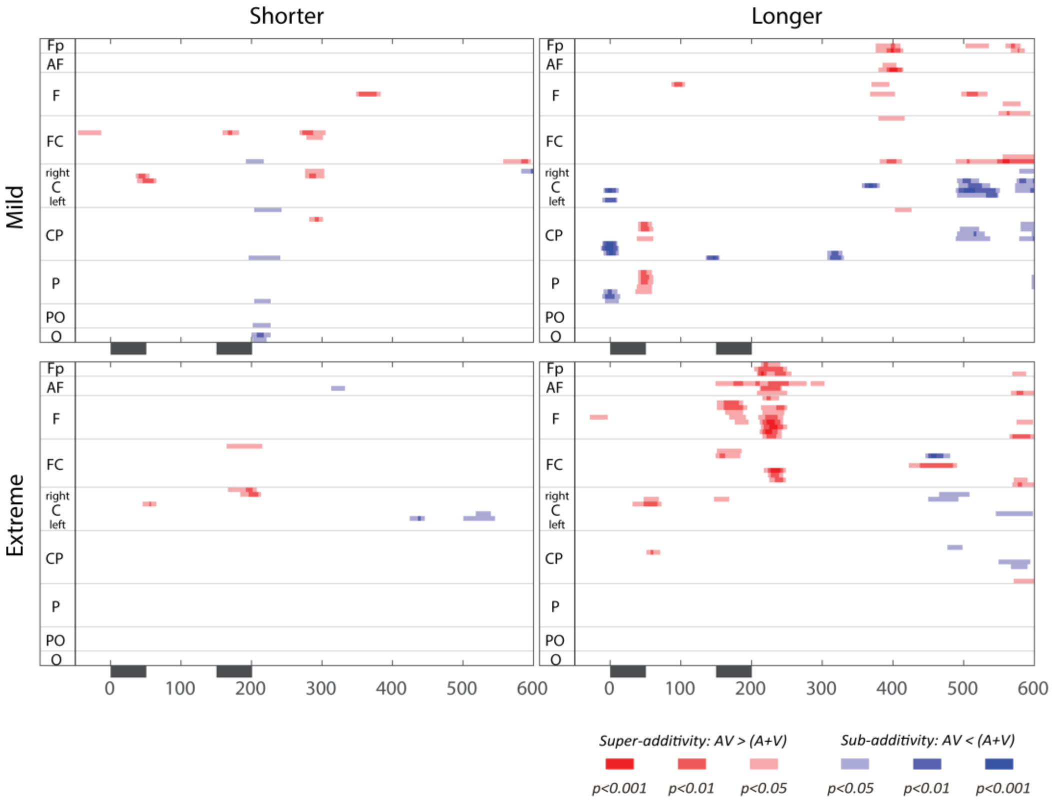


Fig. A1. Spatiotemporal profiles of audiovisual interactions in time interval perception experiment (Experiment 2). AV-A and V responses were compared with the help of the point-wise running *t*-test (two-tailed, *P* <0,05) whether the super- (red) or sub-additivity (blue) was significant or not and the significant differences were represented in the colors of the red and blue where *P* values were significant. Channel-wise comparisons of AV-A and V and shorter and longer modulation conditions are given in left and right panels, and mild and extreme modulations were in upper and lower panels, respectively. Temporal structures of the visual stimuli are marked by black boxes on the timeline.

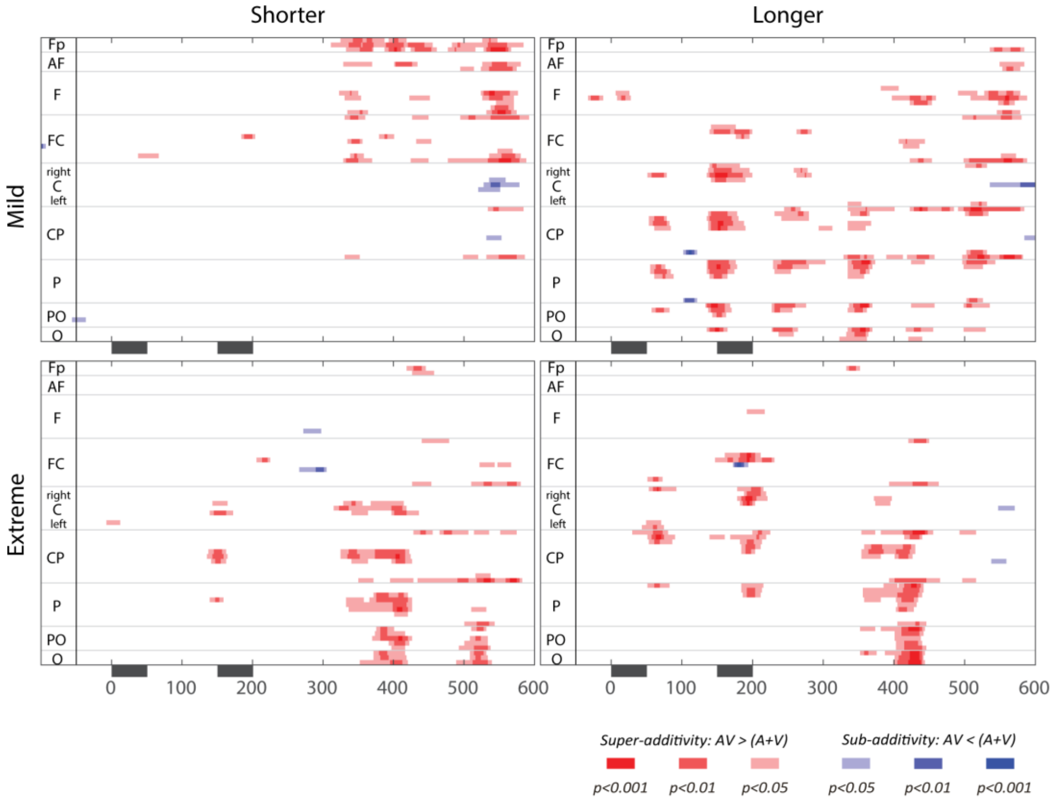


Fig. A2. Spatiotemporal profiles of audiovisual interactions in apparent motion perception (Experiment 3) experiment. Other conventions are the same as in Fig A1.

CV

Utku Kaya, Ph.D.

Aysel Sabuncu Brain Research Center, Bilkent University,
06800, Cankaya, Ankara, Turkey
Ph: +90 535 360 5196 Fax: +90 312 290 3001
E-mail: kayautku@gmail.com

Education

2018 **Ph.D., Cognitive Science,** Middle East Technical University, Ankara, Turkey
Thesis: Neural Mechanisms Underlying Sub-second Crossmodal Time Perception.
Supervisors: Dr. Hulusi Kafaligonul, Dr. Annette Hohenberger

2009 **M.S., Cognitive Science,** Middle East Technical University, Ankara, Turkey
Thesis: An Extended Functionalist Approach to Memetics.
Supervisors: Dr. Ayhan Sol, Dr. Annette Hohenberger

2006 **B.S., Computer Engineering,** Ege University, Izmir, Turkey

Scholarships

2012 National Summer School Support Programme for Turkish Master and PhD Students, The Scientific and Technological Research Council of Turkey (TUBITAK).

2006-2008 Master’s full scholarship. TUBITAK.

Research Interests

Cognitive Neuroscience, Biomedical Imaging, Visual Perception, Time Perception, Multisensory Integration, Modeling of Perceptual and Cognitive Processes

Supports & Research Experience

2016-2017 *Graduate Researcher.* TUBITAK. 1001 Research Grant, Functional Links between Motion Perception and Key Synaptic Targets during Neural Aging (PI: Dr. Hulusi Kafaligonul).
Taken part in designing experimental setup, programming behavioral

experiments (first and second order motion grating), data collection and analyses (fish tracking over video records).

2014-2016 *Graduate Researcher*. TUBITAK. 1001 Research Grant, Neural Mechanisms Underlying Multisensory Perception of Speed (PI: Dr. Hulusi Kafaligonul).
Taken part in designing, programming, and conducting EEG experiments; analyzing and interpreting behavioral and EEG data, writing progress reports and final report.

2013-2014 *Graduate Researcher*. TUBITAK. 3501 Research Grant, Brain Oscillatory Analysis of Short Term Auditory and Visual False Memories (PI: Dr. Tolga Esat Özkurt).
Taken part in designing, programming (Matlab PsychToolbox) and conducting behavioral and EEG experiments (DRM Paradigm, Brain Products 32 channels); analyzing (Matlab FieldTrip Toolbox).

2011-2013 *Graduate Researcher*. TUBITAK. 1001 Research Grant, Early Sensitivity of mono-lingual Turkish Infants for Vowel-harmony and Word Stress (PI: Dr. Annette Hohenberger).
Taken part in writing progress reports and final report; designing (head turn paradigm), programming (E-Prime) and conducting behavioral experiments (with more than 300 babies), analyzing (SPSS) and interpreting data.

Summer Schools & Workshops

2012 New Bulgarian University 19. Cognitive Science Summer School, Sofia, Bulgaria.

2013 Experimental Methods in Language Acquisition Research IX, UiL-OTS, Utrecht University, Netherlands.

Teaching Experience

2018 Spring *Instructor*. Psychology Department, Bilkent University, Ankara, Turkey.
PSYC 310 - Perception, Attention, and Action
PSYC 200 - Cognitive Psychology

2009-2011 *Teaching Assistant***.** Computer Engineering, Atilim University, Ankara, Turkey.

COMPE102 - Computer Programming, COMPE112 - Programming with C, COMPE225 - Object Oriented Programming, COMPE334 - Computer Networks, COMPE431 - Operating Systems, COMPE438 - Java Programming.

2006-2007 *Teaching Assistant*. Informatics Institute, Middle East Technical Uni., Ankara, Turkey.

IS100-Introduction to Information Technologies and Applications.

Publications in Peer-reviewed Journals & Book Chapters

1. **Kaya, U.**, Yildirim, F.Z., & Kafaligonul, H. (2017). The Involvement of Centralized and Distributed Processes in Sub-second Time Interval Adaptation: An ERP Investigation of Apparent Motion. *European Journal of Neuroscience*.
2. **Kaya, U.**, & Kafaligonul, H. (in prep.). Neural mechanisms of crossmodal interactions in the timing of a visual event.
3. **Kaya, U.**, & Kafaligonul, H. (in prep.). Oscillatory mechanisms of auditory effects on a visual time interval and speed perception.
4. Hohenberger, A., Altan, A., **Kaya, U.**, Köksal-Tuncer, Ö., & Avcu, E. (2016). Sensitivity of Turkish infants to vowel harmony: preference shift from familiarity to novelty. In B. Haznedar & F. N. Ketrez (Eds.), *The Acquisition of Turkish in Childhood*, 29-56. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
5. **Kaya, U.**, & Sol, A. (2010). Memetik Modellerin Eliştirel Bir İncelemesi ve Bazı Değerlendirme Ölçütleri. In S. Kibar, S. Aydın Bayram, & A. Sol (Eds.), *Anlam Kavramı Üzerine Yeni Denemeler*, 435-445. Beyoğlu, İstanbul: Legal Kitapevi San. ve Tic. Ltd. Şti.

Conference Proceedings and Presentations

1. Karaduman, A., **Kaya, U.**, Karoglu, E.T., Ergul-Arslan, A., Adams, M.M., Kafaligonul, H. (poster presentation). Motion direction discrimination during neural aging. *Annual Meeting of the Society for Neuroscience*, Washington, DC. November 11-15, 2017.
2. **Kaya, U.**, Yildirim, F. Z., & Kafaligonul, H. Evoked Potentials to Visual Apparent Motion after Auditory and Visual Time Interval Adaptation (Poster Presentation). 40th European Conference on Visual Perception, Berlin, Germany, August 27-31, 2017.
3. Hohenberger, A., **Kaya, U.**, & Altan, A. (2017). Discrimination of Vowel-Harmonic vs Vowel-Disharmonic Words by Monolingual Turkish Infants in the First Year of Life. In *Proceedings of the 41st annual Boston University Conference on Language Development*. Cascadilla Press.
   Poster presentation. The 41st Annual Boston University Conference on Language Development, Boston, NY, November 4-6, 2016.
4. Yildirim, F. Z., **Kaya, U.**, & Kafaligonul, H. (2016). Auditory adaptation alters evoked potentials by visual motion over temporal and frontal regions. *Anatomy: An International Journal of Experimental and Clinical Anatomy*., 10(Supplement 1), 66–67.
   Poster presentation. National Neuroscience Congress 2016, Ankara, Turkey, May 26-29, 2016.
5. Kafaligonul, H., & **Kaya, U.** (2016). Multisensory interactions and perceived timing. *Anatomy: An International Journal of Experimental and Clinical*

Anatomy., 10(Supplement 1), 6.
Oral presentation. National Neuroscience Congress 2016, Ankara, Turkey, May 26-29, 2016.

6. Altan, A., **Kaya, U.**, & Hohenberger, A. (2016). Sensitivity of Turkish Infants to Vowel Harmony in Stem-suffix Sequences: Preference Shift from Familiarity to Novelty. In *Proceedings of the 40th annual Boston University Conference on Language Development*. Cascadilla Press.
Poster presentation. The 40th Annual Boston University Conference on Language Development, Boston, NY, November 13-15, 2016.
7. Kafaligonul, H., & **Kaya, U.** Crossmodal interactions in the timing of a visual event: An EEG study (poster presentation). Society for Neuroscience(SfN) Annual Meeting 2015, Chicago, IL, October 17-21, 2015.
8. **Kaya, U.**, Altan, A., & Hohenberger, A. Early sensitivity of Turkish infants to vowel harmony in stem-suffix sequences (poster presentation). Experimental Methods in Language Acquisition Research (EMLAR) IX, Utrecht, Nederlands, January 30-February 1, 2013.
9. Altan, A., **Kaya, U.**, Köksal-Tuncer, Ö., Avcu, E., & Hohenberger, A. (2013). 6 ve 10 aylık bebeklerin kök ek dizilişlerinde Türkçedeki ünlü uyumuna duyarlılığı. In E. Y. N. Büyükkantarcıoğlu, I. Özyıldırım, I. (Ed.), *27. Ulusal Dilbilim Kurultayı Bildiri kitapçığı*. Kemer, Antalya.
Oral presentation. 27th National Linguistics Congress, Antalya, Turkey, May 2-4, 2013.
10. Fallahzadeh, P., **Kaya, U.**, & Hohenberger, A. Goal-directed Imitation in Four and Five-Year-Old Children (oral presentation). International Conference of Cognitive Science, Tehran, Iran, March 3-5, 2009.
11. **Kaya, U.** A Darwinian Approach to Cultural Change (oral presentation). Darwin Days, Informatics Institute, METU, Ankara, Turkey, November 21, 2009.
12. **Kaya, U.**, & Sol, A. Memetik modellerin değerlendirilmesi için ölçüt geliştirme denemesi (oral presentation). METU Philosophy Days – Meaning Congress, Ankara, Turkey, December 17-19, 2008.

Advising and Mentoring

2015-2018 Sibel Akyüz, PhD Thesis, EEG analyses. Currently Ph.D. student at Neuroscience Program, Bilkent University.

2016-2017 Mert Özkan, Undergraduate senior thesis, EEG analyses. Currently finishing her undergraduate degree at Bilkent University.

2015-2016 Fazilet Zeynep Yıldırım, M.S. Thesis, EEG analyses. Currently Ph.D. student at Institute of Psychology, University of Bern, Switzerland.

2015-2016 Cansu Öğülmüş, M.S. Thesis, EEG analyses. She just finished her M.S. degree from Bilkent University.

Skills & Training

Scientific: Psychophysical Methods, EEG (ERP, Time-frequency), Eye-tracking (Tobii Studio), Statistical Analysis (SPSS, Matlab), Artificial Neural Theory and Applications, Machine Learning Methods, Head Turn Preference Procedure, Praat.

Programming: Matlab (PsychToolbox, Fieldtrip, EEG Lab), Java, C++, C#, SQL.

Software: BrainVision Recorder/Analyzer, SPSS, E-Prime, LateX, MS Office, Adobe Illustrator, Adobe AfterEffects.

OS: Linux, Windows.

Personal Interests

Theatre: Acting, directing, and designing props and sets in many plays in various theatre clubs for more than 15 years.

Cycling: Almost everyday transportation and longer tours.

References

Assist. Prof. Hulusi Kafalıgönül (Advisor, A. S. Brain Research Center)
E-mail: hulusi@bilkent.edu.tr
Telephone: +90 (312) 290 3016
URL: http://hulusi.bilkent.edu.tr/

Assoc. Prof. Annette Hohenberger (Advisor, Cognitive Science Program)
E-mail: hohenber@metu.edu.tr
Telephone: +90 (312) 210 3789
URL: http://ii.metu.edu.tr/staff/hohenber

TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☒

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : KAYA
Adı : UTKU
Bölümü : BİLİŞSEL BİLİMLER

TEZİN ADI (İngilizce) : NEURAL MECHANISMS UNDERLYING SUB-SECOND CROSSMODAL TIME PERCEPTION

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora ☒

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın. ☐
2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullancılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐
3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☒

Yazarın imzası .... Tarih 23.01.2018