AUDIOVISUAL INTERACTIONS IN TIME AND SPATIAL GROUPING PRINCIPLES OF VISION

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

NEUROSCIENCE

By

Cansu Öğülmüş

October 2016

AUDIOVISUAL INTERACTIONS IN TIME AND SPATIAL GROUPING PRINCIPLES OF VISION
By Cansu Öğülmüş
October 2016

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Hacı Hulusi Kafalıgönül(Advisor)

Jennifer Elise Corbett

Annette Edeltraud Hohenberger

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

AUDIOVISUAL INTERACTIONS IN TIME AND SPATIAL GROUPING PRINCIPLES OF VISION

Cansu Öğülmüş
MS in Neuroscience
Advisor: Hacı Hulusi Kafalıgönül
October 2016

Multisensory integration is often studied with intermodal conflict where either visual input dominates and alters the percept of simultaneous auditory input or the other way around. For instance, when put in conflict, visual stimuli can drive the perception of where a sound originates (spatial ventriloquism) [1, 2] whereas auditory stimuli can drive the perception of when visual events occur (temporal ventriloquism) [3, 4, 5] .These interactions make adaptive sense given the auditory system´s superior temporal resolution and the visual system´s superior spatial resolution [6]. Moreover, it was found that temporal ventriloquism can change the perceived speed of visual motion [7]. By taking advantage of this influence of auditory timing on perceived speed, we investigated how audio-visual interactions in time (i.e., temporal ventriloquism) are modulated by the spatial grouping principles of vision. In our experiments, we manipulated spatial proximity, common fate and uniform connectedness between moving flashes. Observers compared the speed of motion between different auditory timing conditions. Our results revealed that auditory influences on perceived speed were significantly modulated by only uniform connectedness. More specifically, we found that auditory effects on vision was significantly less when a horizontal gray connecting bar grouped multiple sequential moving flashes. When horizontally placed moving flashes were grouped with a vertical connecting bar, the degree of auditory influences in time was significantly stronger compared to not grouped (control) and horizontal connecting bar conditions. The effect of auditory clicks on single apparent motion grouped with horizontal connecting bar was smaller relative to the not-grouped condition. In addition, our analysis on EEG activities revealed consistent trends in agreement with the behavioral results. Audiovisual interaction patterns were observed both earlier (around P1) and later (around N1 and P2). Less auditory capture over the horizontal connecting bar condition relative to the vertical connecting bar condition was observed around 50-100 ms (P1) on the frontal and temporal channels and around 200-300 ms (N1) on the frontal, central, temporal and occipital-parietal channels. The larger effect of sound over the single apparent motion without connecting bar condition was observed around 50-100 ms and 200-300 ms on the central channels and 200-300 ms on the occipital-parietal channels. The difference between the individual effects of inner and outer sound conditions was found to be less for the horizontal connecting bar condition relative to the vertical connecting bar condition over the frontal, temporal, central and occipital-parietal channels while less individual effects of the inner and outer sound conditions on the single apparent motion with horizontal connecting bar condition relative to the without connecting bar condition was only observed over the central channels and late time intervals of the occipital-parietal and temporal channels. Overall the individual effects of both sound conditions were consistently similar on the horizontal connecting bar condition (compared to the vertical connecting bar condition) for all ROIs and on the single apparent motion with horizontal connecting bar condition (compared to the no-connecting bar condition) over the central channels. In general, our findings here suggest that temporal ventriloquism effects exist in different spatial grouping conditions of vision, but they can be also modulated by certain intra-modal grouping principles such as uniform connectedness.

Keywords: audiovisual interactions, spatial grouping, speed perception, temporal ventriloquism, ERP, low-frequency brain oscillations, neural mechanisms.

ÖZET

ZAMANSAL DÜZLEMDE GÖRSEL-İŞİTSEL ETKİLEŞİMLER VE GÖRSEL SİSTEMİN UZAMSAL GRUPLAMA PRENSİPLERİ

Cansu Öğülmüş
Nörobilim, Yüksek Lisans
Tez Danışmanı: Hacı Hulusi Kafalıgönül
Ekim 2016

Çoklu duyusal entegrasyon, görsel ya da işitsel uyaranın eşzamanlı bir diğer (işitsel /görsel) uyarandan baskın olup; bu uyaranın algılanış biçimini değiştirdiği modaliteler arası zıtlık yaratılarak çalışılmaktadır. Örneğin, görsel uyaran işitsel uyaranın uzamsal düzlemdeki yerini (uzamsal vantrilokluk) [1, 2]; işitsel uyaran ise görsel uyaranın meydana gelme zamanını (zamansal vantrilokluk) etkileyebilir [3, 4, 5]. Bu etkileşimler, görsel sistemin uzamsal düzlemde işitsel sistemin ise zamansal düzlemde daha duyarlı olmasından kaynaklanmaktadır [6]. Zamansal vantrilokluğun aynı zamanda görsel uyaranın hareket algısını da değiştirdiği görülmüştür. Bu çalışmada, işitsel uyaranın zamanlamasının görsel uyaranın hızı üzerindeki etkisinedayanarak, zamansal düzlemdeki görsel-işitsel etkileşimin (zamansal vantrilokluk) görsel sistemin uzamsal gruplama prensiplerinden nasıl etkilendiğini araştırdık. Yaptığımız deneylerde, zahiri hareketi (görsel uyaran) oluşturan barlar arasındaki yakınlık, ortak hareket ve barların bağlanma (uniform connectedness) parametrelerini değiştirdik. Katılımcılar, farklı işitsel zamanlamalar ile verilen zahiri hareketlerin hızlarını karşılaştırdılar. Deney sonuçlarına göre, barların bağlanması (uniform connectedness) işitsel uyaranın zahiri hareket hızı üzerindeki etkisini önemli ölçüde arttırmaktadır. Yatay gri bir bar, arka arkaya verilen birden fazla barı birleştirip grupladığında; işitsel uyaranın görsel sistem üzerindeki etkisinin en fazla olduğu görülmüştür. Dikey olarak yerleştirilmiş ve yatay hareket eden barlar dikey gri bir bar ile gruplandığında, işitsel uyaranın görsel uyaran üzerindeki etkisi birleştiren gri barın olmadığı ya da birleştiren yatay gri barın olduğu görsel gruplama olmayan durumlara göre daha fazla olduğu görülmüştür. Tek zahiri hareketin barları yatay gri bir barla birleştirildiğinde, işitsel uyaranın etkisinin daha az olduğu görülmüştür. Bununla beraber, EEG aktiviteleri üzerinde yapılan analizler davranışsal deneylerdeki sonuçlarla uyumlu bir eğilim ortaya çıkarmıştır. Görsel-işitsel entegrasyon ağı hem erken (P1) hem de geç (N1 ve P2) zaman aralıklarında gözlenmiştir. Yatay barla gruplanan zahiri hareketlerin dikey barla gruplananlara göre sesten daha az etkilenmesi temporal ve frontal kanallarda 50-100 ms aralığında, temporal, frontal, santral ve oksipital-parietal kanallarda ise 200-300 ms aralığında görülmüştür. Tek zahiri hareket durumlarında ise yatay gri barla gruplanıldığında sesten daha az etkilenme durumu santral kanallarda 50-100 ms ve 200-300 ms aralıklarında ve oksipital-parietal kanallarda sadece 200-300 ms zaman aralığında gözlenmiştir. Bununla birlikte frontal, temporal, santral ve oksipital-parietal kanallardaki 50-100 ms, 200-300 ms ve 300-400 ms zaman aralıklarında hızlı ve yavaş seslerin bağımsız etkilerinin, dikey barla bağlanmış zahiri hareketlerde yatay barla bağlanmış zahiri hareketlere göre daha fazla olduğu görülmüştür. Ayrıca hızlı ve yavaş seslerin bağımsız etkileri arasındaki fark yatay barla bağlanmış tek zahiri hareket gösterildiğinde barla bağlanmamış tek zahiri hareket gösterildiği duruma göre santral kanallarda hem erken hem geç zaman aralıklarında, oksipital-parietal ve temporal kanallarda ise sadece geç zaman aralıklarında daha az bulunmuştur. Farklı ses koşullarının etkisi yatay barla bağlanmış zahiri harekete/hareketlere göre dikey barla bağlanmış ya da bağlanmamış zahiri harekete/hareketlerde yakın bulunmuştur. Sonuç olarak elimizdeki verilere göre, görsel ve işitsel modaliteler arası gruplama yapılmadan önce görsel modalite içinde bir gruplama yapılırsa, işitsel uyaran görsel uyaranın zaman algısını değiştirebilir.

Anahtar sözcükler: Çoklu duyusal entegrasyon, uzamsal gruplama prensipleri, hareket algısı, zamansal vantrilokluk, ERP, düşük frekanslı beyin salınımları, sinirsel mekanizmalar.

Acknowledgement

I would first like to express my deepest sense of gratitude to my advisor Asst. Prof. Dr. H. Hulusi Kafalıgönül for his continuous support, patience, motivation and enthusiasm throughout the course of this thesis. I thank him for the guidance and great effort he put into training me in the scientific field. Besides, I would like to thank Asst. Prof. Dr. Jennifer Corbett and Assoc. Prof. Dr. Annette Hohenberger for their interest in my work and for their insightful comments.

I must express to my sincerest thanks to my dearest friends and labmates Utku Kaya for his support during EEG recording&analyses; Koray Ertan for his assistance on programming, Merve Karacaoğlu for the help she provided on behavioral experiments, Buse M. Ürgen, Zahide Pamir, Ayşenur Karaduman, Pınar Demirayak and F. Zeynep Yıldırım for their valuable comments on my work and the emotional support in the course of my masters study. I would like to thank the entire UMRAM team for the friendly and cooperative atmosphere.

Finally, I am grateful to my parents Birgül and Cantaş Öğülmüş and my sister F. Ezgi Öğülmüş for the patience, unconditional support and encouragement they provided throughout my life and during the thesis process. I am thankful to my beloved friends Okan Doğan, Uygar Altınok, Ö. Sıla Gündüz, Meltem Okan and D. Ekin Demirci for the unyielding spiritual support.

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

Contents

List of Figures

Chapter 1

Introduction

1.1 Audiovisual Interactions in Time

Perception of the world depends on the interaction of simultaneous information from multiple sensory sources. Brain gathers all information from available sensory inputs and integrates them in order to generate a final percept. However, different sensory systems occasionally provide conflicting information. In these circumstances, final perception is based on the sense which is dominating over the other ones. For instance, in the spatial domain, the dominant sense is vision relative to audition. Thus, in case of simultaneous auditory and visual presentation, information from the auditory system about the spatial location of a stimulus can be modulated by the conflicting information from the visual system [11]. This phenomenon is called spatial ventriloquism and it is propounded by the example of McGurk effect. In this illusion, observer sees the speaker saying 'ga', but at the same time hears the sound 'ba'. The final perception of the sound is influenced by the inconsistent input from the visual and the auditory systems and observer perceives a different sound, such as 'da' [12]. Audiovisual interaction in the spatial domain occurs due to the dominance of visual system, in other words "what is being heard is influenced by what is being seen". According to the

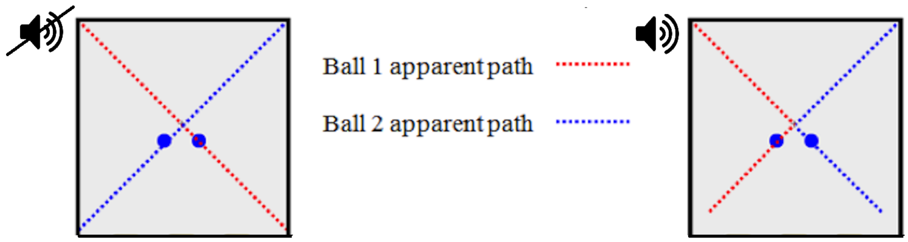

Figure 1.1: Motion-Bounce Illusion

discriminability-dependent-weighting hypothesis [4] the reason for spatial ventriloquism is that vision provides more precise information in the spatial domain and therefore it is more dominant relative to auditory information. It has been suggested that spatial perception depends on the modality with better quality [13]. This capturing effect of the visual system on the auditory system rapidly decreases as the attention is directed towards the auditory stimulus or when the visual stimulus becomes more ambiguous. On the other hand, the visual system fails to provide precise information in the temporal domain. The auditory system has better temporal resolution and hence the information about the time of occurrence and duration are more viable in the simultaneous auditory source [6, 14]. Decisions regarding the time of the audiovisual event are dominated by the auditory input through modulating the information from the visual system. This audiovisual interaction in the temporal domain is called temporal ventriloquism [4]. One of the well-known examples of temporal ventriloquism is the Motion-Bounce Illusion(Figure 1.1). It is basically the perception of bounce when an auditory stimulus is presented at the contact point of two visual stimuli that are moving towards each other [15]. Without the auditory stimulus, observer perceives that each ball follows two distinct pathways (demonstrated in red and green dashed lines) from the upper left corner to the lower right corner and from the upper right corner to the lower left corner. When a static click is introduced at the contact point of two balls, observer perceives that the balls knocks against each other at the contact point and bounce away. Thus the pathway that each ball follow is changed with the modulatory effect of the click.

1.2 Apparent Motion and Auditory Timing

Apparent motion is basically a motion illusion generated by the rapid sequential occurrence of two flashes at different spatial locations [8]. Previous studies on audiovisual integration mostly used apparent motion as the visual stimuli [8, 16]. Visual flashes were perceived to move over the most direct and shortest path rather than being disappeared and a second flash appeared in another location [10]. There are three essential aspects of apparent motion. First one is the exposure time of the flashes. The flashes should be presented long enough to be perceived. Optimal exposure time is determined as 50 ms. Other one is the spatial distance between the flashes. If the distance is too long, observer does not experience a motion, instead two irrelevant flashes at different spatial locations. Third one is the temporal distance between flashes. Previous research revealed specific range of inter-stimulus interval (ISI), time interval between first and second flashes, where motion illusion can be experienced. If the ISI is too short, less than 50 ms, two flashes seem to occur simultaneously. If the ISI is too long, no motion is experienced. Observers mostly perceive two irrelevant flashes appear close to each other but not related [8]. As reported by Strybel and colleagues [17], the range of ISIs for a clear apparent motion is between 50 and 150 ms. Static sounds (clicks) are mostly used for auditory stimulation in apparent motion studies. Several studies suggested that apparent motion is modulated remarkably by the sounds presented temporally to the visual stimuli [7, 8, 16].

Auditory timing, namely the time the clicks are introduced, determines the extent of influence that the clicks have on the visual stimuli. Study conducted by Kafaligonul and Stoner [7] revealed that clicks introduced between the flashes in time can lead to a reduction in perceived ISI between the flashes. It was revealed that when the ISI between the auditory clicks was smaller than the ISI between the flashes, apparent motion was perceived to be faster than it physically was. In other words, these clicks contracted the time interval between the flashes. Besides, when the ISI between the flashes was smaller than the ISI between the clicks, motion was perceived slower due to the modulatory effect of auditory timing on ISI between the flashes [7]. Similar findings were obtained by Morein-Zamir and colleagues [4]. It was found that the clicks with an ISI smaller than the one between the flashes pulled the flashes closer in time and eventually observers performed worse in reporting which of the unimodal cues was presented first in a cross-modal stimulus (temporal order judgment task). A study by Getzmann [8] was in compliance with these findings and further indicated that sounds presented between the first and second flash resulted in a continuous apparent motion even at larger ISIs between these flashes. It was previously proposed by Shepard and colleagues [10] that large ISI between flashes disrupted clear continuous apparent motion and instead gave a percept of two irrelevant flashes. However, according to Getzmann [8], clicks introduced between flashes had a shrinking effect on apparent motion even at larger ISI values.

Nonetheless, the temporal ventriloquism effect is not robust if the time interval between flashes and clicks are close to 0 or more than 250 ms. If the time interval is close to 0, the flash and the click are perceived simultaneously and hence the audiovisual effect is not observed. Likewise, when the interval between the flash and the click is more than 250 ms, the click becomes irrelevant to the flash and no longer has an effect on the flash [4]. The effect of each click was investigated separately in order to see whether the click preceding the first flash or the click following the second flash was responsible for the shift in time interval in apparent motion (Figure 1.2). Clicks preceding and following apparent motion were compared to the clicks presented simultaneous to the flashes. It was found that only the click following the second flash contributed to the visual task performance [4].

Second click was found to be shifting the time interval between the flashes away while no modulatory effect found for preceding and simultaneous clicks. Proposed explanation for the asymmetry between proceeding and following clicks is that the auditory processing is faster than the visual processing. Preceding click is processed faster than the first flash but second flash is still in processing stage when following click is introduced. By this way, perception of second flash and following click is interfered; resulting in an audiovisual interaction which affected observers' performance [4]. It can be concluded that two clicks accompanying

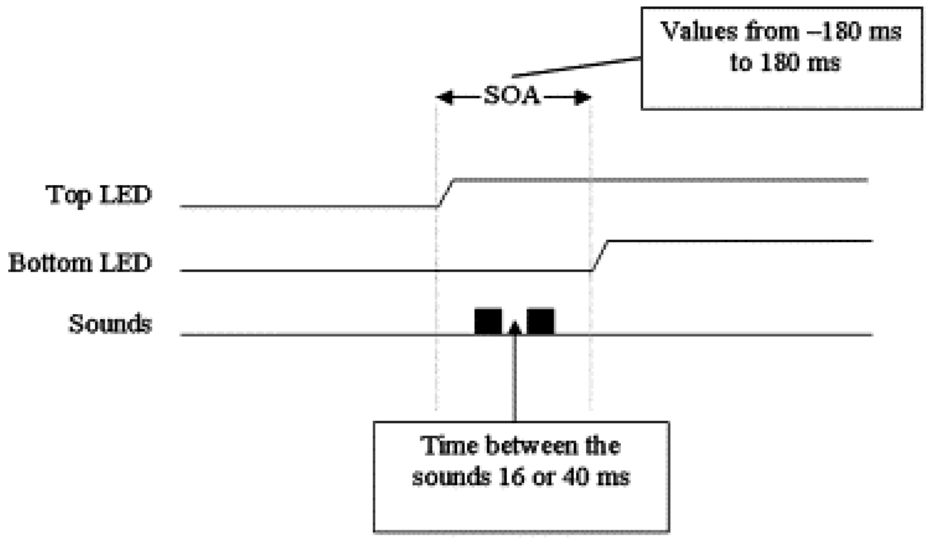

Figure 1.2: Schematic illustration of experimental design used by Morein-Zamir et al. [4]

apparent motion should be located before the first and after the second flash in order to pull these flashes apart in time and after the first and before the second flash in order to pull these flashes closer in time. In the first condition, second click following the second flash and in the second condition first click following the first flash should be taken into consideration.

1.3 Grouping versus Cross-modal Interaction

Audiovisual integration is an inter-modal pairing where inputs from the auditory system and visual system interacts and contributes to the final perception. Previous studies suggest that certain properties are needed for audiovisual interaction. One of these properties is the number of items from each modality. Morein-Zamir and colleagues [4] observed the quality of audiovisual integration when the auditory stimuli are less in number compared to the visual stimuli. More specifically, the effect of a single click centrally located to the apparent motion was investigated. No significant effect of the click, introduced in the middle of the apparent motion, indicated the importance of the number of clicks needed to be introduced with the flashes [4]. A remarkable effect of one modality over another was only observed when the flashes and clicks were equal in number. Thus, a single sound was too weak to shift the time of occurrence of the apparent motion since single sound could not bind to the two sequential flashes of apparent motion. This conclusion was in accordance with the study by Shams and colleagues [18] which suggests that multiple clicks create the illusion of multiple flashes even though there is only one flash presented. It basically indicates that multiple clicks manipulate the quantity of the visual stimulus. Illusory flashes are perceived as a consequence of a need for a pairing between the visual and auditory modalities. These two studies supported the assumption of unity. In brief, the assumption of unity suggests that two or more sensory inputs that share common properties, such as sharing the same temporal dimension, spatial dimension, number, semantic content and so on, are more likely to be originated from the same multisensory source, rather than two or more unimodal sources [19, 20]. Complementary findings of Getzmann [8] demonstrated that a single click introduced in the midway of apparent motion has a modulatory effect on both flashes when the time interval between the flashes are more than 100 ms. The aim of the study was to investigate whether the number of input from each modality mattered in audiovisual interaction. The double click condition, with a constant ISI clicks value of 16 ms, the two intervening clicks condition, with variable click ISIs, and the single click condition were compared to each other (Figure 1.3). It was found that when the time interval between the flashes was less than 100 ms, there were no significant differences across conditions. When the ISI between flashes was 150-250 ms, the double click and the single click conditions were found to be facilitating the apparent motion more than the two intervening clicks condition did. It indicated that the single click introduced in the midway of the apparent motion has a manipulative effect even the ISI between flashes is high. A similar effect was also observed for the double clicks condition too, but the reason may be that observers perceived the double clicks as a single click due to the small ISI value.

One explanation for the observed effect of single click may be that apparent

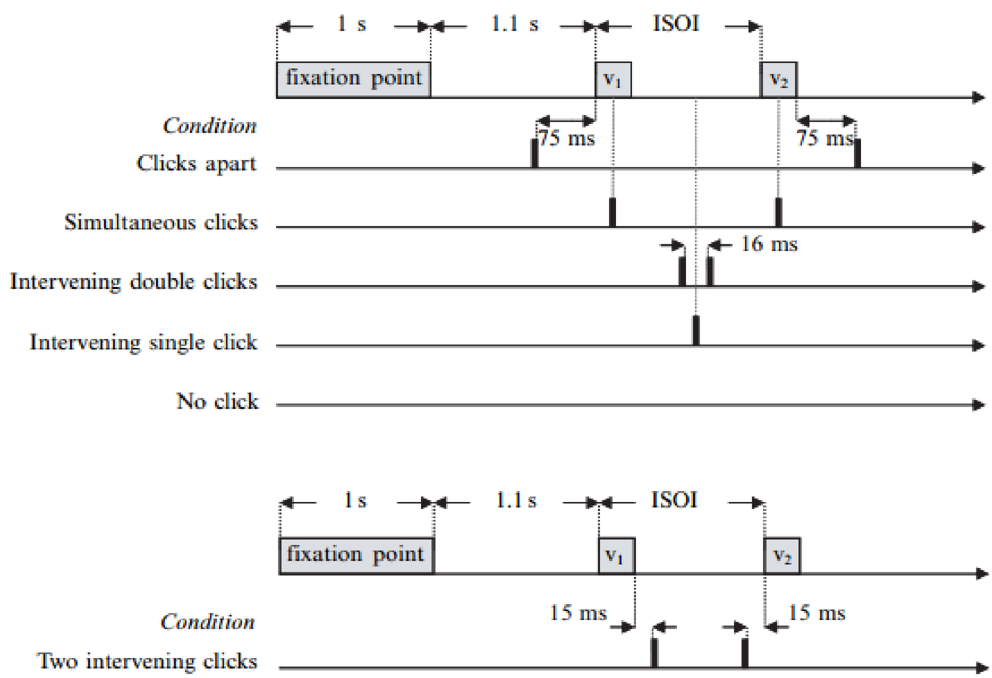

Figure 1.3: Experimental design by Getzmann [8]

motion is perceived as a single object moving from one location to another. Thus, a single click would be sufficient to pair with a single flash moving [8]. From this point of view, even though the findings seem to be contradictory to the previous studies, it is in accordance with the assumption of unity. One click and one continuous flash are equal in number. Another explanation might be that the time that clicks are introduced. The ISI between clicks and flashes carry more importance than the number of clicks introduced. In this case, the single click was presented in the midway of the apparent motion and it facilitated the continuity better than the clicks that were presented closer to each flash. Alternatively, intervening clicks binding to each flash and pulling them away may be disrupting the continuity of apparent motion. With the increase in ISI between flashes, two intervening clicks were paired with two flashes and generated distinct audiovisual events, which led to a decrease in continuity of apparent motion [8]. These two conclusions are both in accordance and complementary to the previous studies.

The assumption of unity in inter-modal interaction suggests that elements from each modality need to be equal in number. This way, each element can match to one other on the other modality and interact. If the elements from each modality are not equal in number, intra-group pairing takes place in order to group the elements of each modality and prepare them for the inter-group pairing. Grouping in the visual and auditory systems operate similarly. Shipley and colleagues [21] reported that the rapidly alternating visual stream can be modulated by the simultaneous fluttering sound stream. Bregman [22] suggested that stream segregation in an auditory stream consisting of multiple sounds was possible in faster rates of alternation between the sounds when the frequency difference between the sounds was high. It basically implies that if the frequency difference between the tones is small; one does not recognize the alternation between the sounds and thus cannot segregate the sound sequences into two. On the contrary, if the tones differ from each other at a great deal, observers perceive the sequence of sounds as two different streams even at fast alteration rates. A similar grouping principle was observed also in the visual system and called visual stream segregation [23]. Series of lights were presented sequentially in lower and higher area of the visual field in an alternating pattern, (e.g.: high-low-high-low-high-low so on.) If the alternating speed was slow, observer perceived the illusion of a flash moving up and down. However, if the sequence was presented at a faster rate, observer perceived two different streams of flashes on their shortest paths, (e.g.: high-high-high or low-low-low). It can be concluded that the auditory and visual grouping is alike.

1.4 Gestalt Grouping Principles

Gestalt grouping principles have been mostly studied in the spatial domain. In essence, gestalt psychologists suggested an organization of scene according to certain physical properties of the visual inputs [24]. These principles include proximity, similarity, good continuation, common fate and closure. Grouping by proximity suggests that objects that are closer to each other in the spatial domain tend to be grouped together (Figure 1.4B). Similarity, on the other hand, is about being physically similar such as in shape and color (Figure 1.4A). If two objects look similar, they are going to be grouped together

Both visual grouping principles have been widely used to understand basic grouping mechanisms in the visual system. Research by Chen [25] investigated whether the proximity or the similarity grouping principle was more fundamental in the visual system. In a design where proximity and similarity provided conflicting cues about grouping of items that were placed in rows and columns, observers tended to group the items by proximity if all the objects were presented only for a short period of time. However, if the scene was presented for a longer time, observers tended to group them by similarity (i.e., geometrical properties) instead. It was concluded that proximity proximity was more fundamental (or primitive) in grouping of visual scene.

In addition to these grouping principles, good continuation principle suggests grouping by illusory connection of the contours in different spatial regions into one as they generate one smooth figure (Figure 1.4C). Common fate principle, on the other hand, is grouping the elements that move to the same direction around the same time (Figure 1.4D) [26]. Proximity principle was found to be dominant over grouping by good continuation principle as well. Kurylo and colleagues [27] reported shorter response times for objects grouped by spatial proximity compared to the grouping by good continuation. Overall, the literature investigating the effectiveness of these four visual grouping principles indicated the dominance of the proximity principle over the rest. Shorter response times for grouping objects that are close to each other demonstrated that grouping by proximity operated with less effort.

1.5 Uniform Connectedness

According to the literature mentioned above, the spatial proximity principle seems to be the most dominant and fundamental spatial grouping strategy, since it groups visual stimuli presented even for a short period of time. Nevertheless, Palmer and Rock [9] argued that what is more fundamental than grouping by

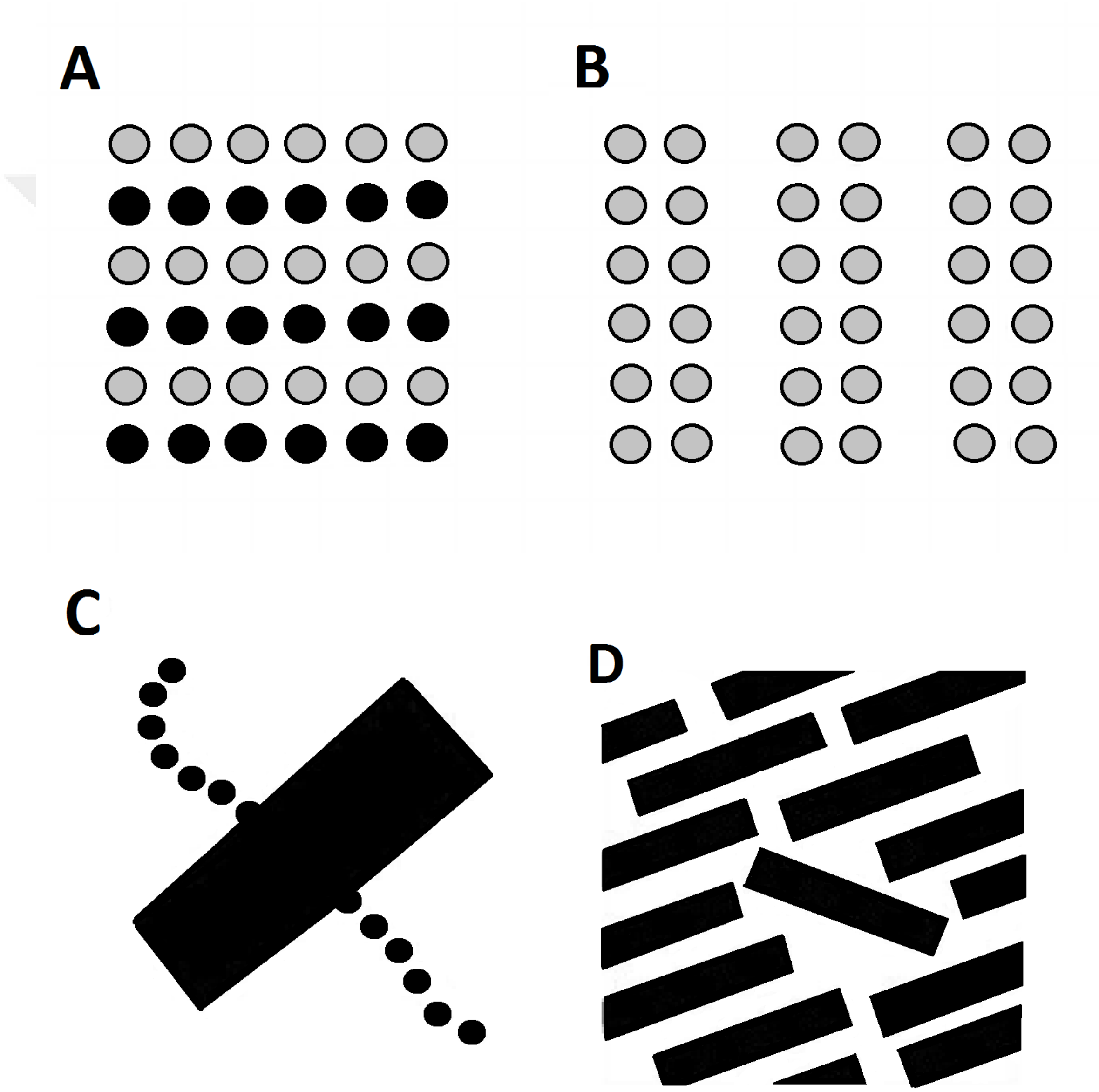

Figure 1.4: Gestalt spatial grouping by similarity (A), proximity (B), good continuation (C) and common fate (D) principles

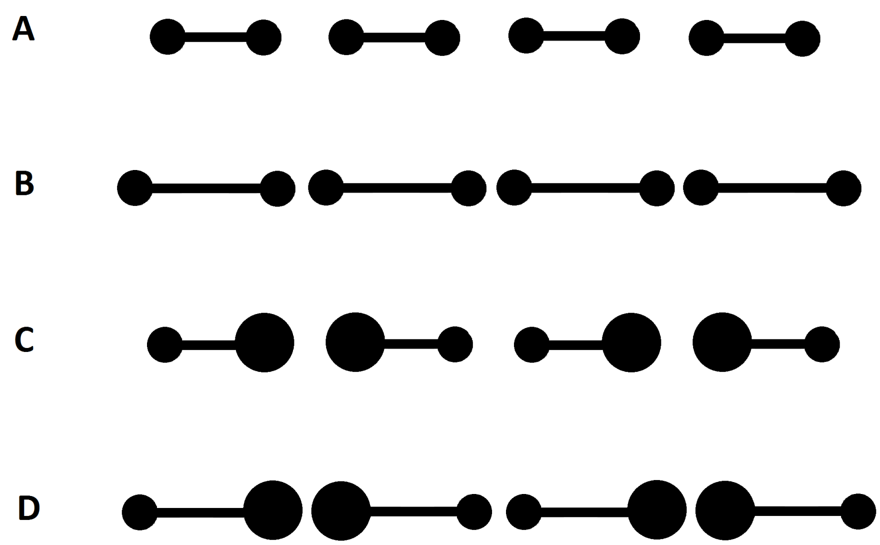

Figure 1.5: Two dots connected with uniform connectedness grouping principle (A) are found to be dominating grouping by proximity (B), size (C) and both proximity and size (D). Taken from Palmer et al. [9]

proximity is grouping by uniform connectedness principle (UC). If two or more objects are physically connected to each other, they are represented at the retina as one and therefore they enter the visual system as a single object. Physically connected objects are hence primitively grouped and more dominant than any other spatial grouping strategy, including proximity, similarity or both (Figure 1.5).

Following study by Han and colleagues [26] compared the reaction time differences between grouping by similarity, proximity and uniform connectedness principles. Conditions in which uniform connectedness principle was applied to the visual stimuli that were already grouped with either similarity grouping principle or proximity grouping principle were formed. For instance, eight circles placed within squares indicated similarity grouping principle. If these circles were physically connected by solid lines, uniform connectedness principle was applied to visual stimuli that has been already grouped with similarity principle

(UC+ similarity condition). Similar example was true for condition with proximity grouping principle, except the presence of squares and circles. Instead, all the visual stimuli were circles but they were grouped by their physical distance and connecting solid lines (UC+ proximity condition). Observed faster reaction times in UC + similarity condition compared to the similarity only condition revealed that uniform connectedness facilitated the grouping by similarity principle. However, similar effect was not observed for the proximity condition. The reaction time difference between the circles grouped with physical distance and the UC + proximity condition was not significantly different from each other [26]. It was concluded that while the grouping by the uniform connectedness principle dominated over grouping by the similarity principle, grouping by the proximity principle was as strong as grouping by the uniform connectedness principle. One proposed explanation for the similar recognition efficiency of grouping by the proximity and uniform connectedness principles is that they both concern low spatial frequency channels. Ginsburg [28] indicated that the circles and the lines connecting the circles are filtered and only the low spatial frequencies are able to pass. Since underlying shape is kept in both grouping principles, their recognition efficiency should be the same. Another explanation for the similar effects of grouping by the uniform connectedness and by the proximity principles is that global forms are generated by local elements forming edge boundaries in both grouping condition. In other words, virtual boundaries are formed around the shapes generated in both proximity and uniform connectedness condition. Hence, even there are gaps between the elements that form the shape, edge detectors still generate boundaries. Nevertheless, these boundaries are not formed when the spatial distance between the elements of the shape is large or misaligned [29].

1.6 Path Guided Apparent Motion

Considering uniform connectedness is an important visual grouping principle, apparent motion studies with uniform connectedness principle employ path guided apparent motion. Basically, it is a ghost-like path between the first and second

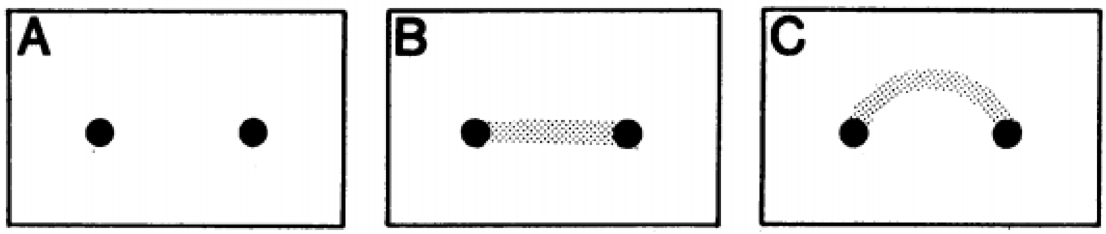

Figure 1.6: Classical apparent motion moving back and forth over the shortest path (A), path guided apparent motions providing an imitation of rapid motion (B and C) taken from Shepard et al. [10]

sequential flashes of a regular apparent motion. In a real motion, sensory receptors are stimulated on this blurry path. Typically, regular apparent motion is generated by observer perceiving two sequential flashes as moving from the most direct path (Figure 1.6A). Shepard and colleagues [10] suggested that brief presentation of a faint connecting band between the flashes would stimulate the receptors that are known to be stimulated during a real motion (Figure 1.6B). It was concluded that experienced motion was weaker in the absence of path. It was further investigated whether the flashes seem to move on the path even though the path is curved (Figure 1.6C). According to Korte´s law, apparent motion occurs when the flash prefers the most direct path to the arriving location; not a curved, longer path. It was revealed that quality of apparent motion depended on length of the curved path rather than physical distance between the flashes [10].

1.7 Levels of Processing and Neural Correlates

The intra-modal grouping was previously mentioned to be prior to the inter-modal grouping. Visual stimuli and auditory stimuli are grouped within their modalities; then the group of elements from one modality binds to the group of elements from the other. According to Schröger and Widmann [30], audiovisual interaction is not at an entry level of processing since it follows sensory processing of the items from each modality. It is rather stated that audiovisual interaction occurs at intermediate level of processing. More specifically, before the late, ‟low level ″motor stages where observer is prepared to generate an action and later than early, ‟sensory specific ″stages where the elements from each modality are processed. Besides, EEG data indicated that the auditory and visual stimuli were independently processed for a short amount of time. Thereafter an integrated processing was observed in the first 200 ms post-stimulus onset [30]. It was concluded that the integration occurred after the sensory processing. Behavioral data indicated the existence of a combined processing of the audiovisual information since the reaction times for the bimodal stimuli were found to be shorter compared to the unimodal stimuli [30]. Miller and colleagues [31] suggested two possible processes responsible for the processing of audiovisual stimulation. The auditory and visual information in the bimodal stimulus are processed separately and their summation builds up the final output (independent co-activation hypothesis) or the visual and auditory processing are not independent; instead the presence or absence of the information in one modality influence the other modality (interactive co-activation). EEG study conducted by Giard and colleagues [32] demonstrated evidence for the interactive co-activation hypothesis. In an object identification task that participants were presented both unimodal and bimodal stimuli, cross-modal interaction patterns was observed on both sensory-specific visual and auditory regions and non-specific areas (right fronto-temporal cortices) between 40-200 ms post-stimulus onset. Bonath and colleagues [33], investigated the neural correlates of spatial ventriloquism and indicated laterally biased cortical activity in the N260 component. Furthermore, it is stated that the N260 component is colocalized with BOLD response in posterior/medial regions of the auditory cortex [33]. It was stated that audiovisual interaction occurred around 260 ms after stimulus onset. However, more recent studies indicated audiovisual patterns at earlier latencies. First positive peaks around 100 ms was observed in response to the audiovisual stimulus [34]. Moreover, first negative peak around 160-180 ms was also considered as an audiovisual pattern [35]. A study by Freeman and Driver [16] reported the contribution of the higher-order association areas on audiovisual interaction. However, the reason for the role of higher order association areas might be the long-range apparent motion stimuli used by

Freeman and Driver [16]. A more recent study by Kafaligonul and Stoner [7] indicated that audiovisual interaction can occur at relatively lower stage in the visual processing: area MT/V5. Audiovisual properties of area MT/V5 were revealed by a TMS study of Bueti and colleagues [36] where subjects with temporarily impaired area MT/V5 were measured on the temporal intervals discrimination task. It was indicated that the motion sensitive regions such as MT/V5 might be involved in temporal discrimination. Interestingly, this study revealed that the area MT/V5 that is known to be involved in motion perception is also important for temporal processing. A possible explanation proposed for the same brain area for both spatial and temporal processing is that human use temporal mechanisms for carrying out actions in space [36]. There are also neurons responding to audiovisual stimuli in prefrontal cortex, periarcuate cortex [37] and orbitofrontal cortex [38]. After all, literature supports the involvement of area MT/V5 and higher cortical areas on processing of the audiovisual stimuli. Some studies in the literature supported the additive view proposed by Stein and colleagues [39]. It proposed that the audiovisual stimuli consisted of sensory processing of the visual and auditory information. Thus, audiovisual interaction is the linear summation of its auditory and visual elements. However, Giard and Peronnet [32] reported that in some cases, visually-evoked N1 decreased its amplitude in response to the audiovisual stimuli, and this decrease is called sub-additive interaction, or cross-modal depression. The proposed explanation for the reduced visually evoked N1 component was that the presence of auditory stimulus decreased the need to engage attention to the visual stimulus [32]. Sub-additive interaction indicates that audiovisual interaction is not a summation of the unimodal components; instead it is a distinct process. Similarly, when the response to the audiovisual stimulus is larger than the sum of response to the unimodal auditory and visual stimuli, nonlinear multisensory effect was observed and it is called supra-additive interaction [40].

1.8 Specific Aims

Current study aims to observe the auditory capture of visual stimuli in the temporal domain by using apparent motion stimuli under different gestalt grouping principles. By behavioral experiments, audiovisual interaction with multiple apparent motions and two static sounds was investigated. Three different spatial organizations were created with gestalt spatial proximity, common fate and uniform connectedness principles and the differences in the auditory capture on these spatial organizations were observed. Proximity principle was applied by defining the spatial distances between the two vertically placed apparent motions. The apparent motion stimuli that are spatially closer to each other are expected to be captured by the accompanying clicks more than the apparent motion stimuli that are in distance. The common fate principle was applied by assigning the apparent motion on the bottom to move in left -right direction and the apparent motion on the top to move in up-down direction. The apparent motion stimuli that move in the same direction (control condition) are expected to be grouped together and the effect of sound is expected to be less compared to the apparent motion stimuli moving to different directions. The uniform connectedness principle was applied with the vertical and horizontal fainted connecting bars. When these connecting bars are placed between the first and second flashes of each apparent motion, multiple apparent motions are expected to group as two/three distinct objects moving in left-right direction. Since the inter-modal pairing between three apparent motions and two static clicks would be harder, apparent motions grouped with horizontal connecting bars are expected to decrease the modulatory effect of the auditory stimuli. On the other hand, when the fainted connecting bars are placed in between the first and second flashes, multiple apparent motions are expected to group as one bigger apparent motion and captured by the accompanying sounds.Moreover, in case of single apparent motion stimuli, the horizontal connecting bars, connecting the two sequential flashes, is expected to create a clear apparent motion illusion. The temporal occurrence of the single apparent motion with horizontal connecting bar condition expected to be modulated by the static sounds less than the single apparent motion without connecting bar condition.

Furthermore, on EEG data, for all visual grouping conditions, the voltage differences associated with the sound conditions are expected around 100-200 ms post-stimulus onset. In the bimodal conditions with multiple apparent motions grouped with horizontal connecting bar, voltage difference due to the sound conditions are expected to be less compared to the multiple apparent motions with vertical connecting bar and without connecting bar. Similarly, in the bimodal conditions with single apparent motion, voltage difference due to the sound conditions are expected to be less when horizontal gray connecting bar is introduced between the flashes. Moreover, voltage differences associated with inner and outer sound on all bimodal conditions are expected to occur on occipital, parietal and temporal regions.

It is hypothesized that intra-modal grouping occurs prior to pairing between modalities. It is suggested that, audiovisual interaction occurs only if stimulus for each modality engages within modality mechanisms, and then bind to (group with) the elements of other modality.

Chapter 2

Behavioral Experiments: Method and Results

Audiovisual interaction studies employing visual motion stimulus mostly used apparent motion stimuli with both short and longer ISI values [7, 16, 17, 41]. However, it was shown that shorter temporal interval between two flashes results in a more clear motion perception [16]. Blurry connecting bar introduced by Shepard and colleagues [10] has been suggested to create a continuous apparent motion when placed between the two flashes of apparent motions. Palmer and Rock suggested that uniform connectedness is the most fundamental spatial grouping principle in the visual system [9].

The current study employed the fainted gray connecting bar introduced in Shepard's study to group multiple apparent motion stimuli such that the connecting lines of uniform connectedness principle group individual objects. The spatial organizations generated with these connecting bars compared to the multiple apparent motions without connecting bars. Auditory stimuli were defined as two set of sequential static sounds with two different ISI values since Kafaligonul and colleagues [7] concluded that static sounds introduced either before the first and after the second flash (outer) or after the first and before the second flash (inner) can capture the short-range apparent motion stimuli.

All participants underwent a training session consisting of random presentation of all visual grouping conditions without the accompany of the clicks. The participants were asked to state the number of visual stimuli they perceived via a standard keyboard. The training sessions lasted approximately 40 minutes. The training session was implemented prior to each experimental session.

First experiment employed two horizontally placed apparent motions. On each trial, these apparent motions were grouped with one of the connecting bar conditions (vertical, horizontal, aligned), spatial proximity conditions (three vertical distance between apparent motions) and common fate conditions (direction of movement). Experiment 2 investigated the effect of sounds on horizontally placed three apparent motions grouped with vertical connecting bar and horizontal connecting bar. Experiment 3 examined the capture of sounds on single apparent motion grouped with horizontal connecting bar and without connecting bar. Details of each experiment were discussed in the following section below. Data from each experiment was obtained in different sessions.

2.1 Experiment 1- Two Apparent Motions under different Gestalt Grouping Principles

Previous studies suggested that intra-modal grouping may have an effect on following cross-modal interaction [4, 23, 42]. Hence, different spatial organizations of visual stimuli are expected to determine the quality of auditory capture. This experiment was designed to investigate the differences between auditory capture on apparent motions grouped with three different gestalt spatial grouping principles: spatial proximity, common fate and uniform connectedness.

Subjects

Participants were 14 volunteers (4 female, 10 male) from Bilkent University and all were naïve to the purpose of the experiment. The participants had normal or corrected-to-normal visual acuity and none was suffered from any kind of auditory dysfunction. Prior to each experimental session, all participants underwent training. After the training phase, they completed one experimental session including all visual grouping and auditory conditions. Participants gave informed consent of participation and all procedures were in accordance with international standards (WMA Declaration of Helsinki) and approved by Ankara University Ethics Committee.

Apparatus

Visual stimuli were presented on a 20 CRT monitor (HP P1230, 1280 x 1024 pixel resolution and 100 Hz refresh rate) at a viewing distance of 57 cm. Auditory stimuli was introduced via Sennheiser HD518 headphones.

The participants were seated in a dark, sound-attenuated psychophysics room. A head and chinrest was used to keep the distance between the eye and display stable across participants.

Stimuli and Procedure

Fixation point was a small bright red circle (0.3°diameter) at the center of the display. The visual stimuli was two vertically placed simultaneous apparent motions presented on a black background. Each apparent motion consisted of horizontally placed sequential flashes (0.3 x 1.2 degree, with a luminance of 35 cd/m^2) from one location to another (0.8°spatial distance between the flashes). These two simultaneous apparent motions were always presented on the right hemi-field of the display. Because the spatial location of the apparent motions (left, right, up or down) did not have an impact on the task. Moreover, when the visual stimuli were randomly presented at different left and right hemi-fields, participants were highly distacted. Also we did not want the expectation of spatial location to have an influence on the results. Thus, participants were asked to pay attention to the right hemi-field of the display throughout the experimental session.

Two different spatial organizations were generated by introducing fainted gray connecting bars, (with a luminance of 35 cd/m^2) between the flashes of these two apparent motions. Horizontal connecting bar introduced between the two sequential flashes of each apparent motion formed the first grouping by UC condition called *horizontal connecting bar condition* and created a perception of two distinct simultaneous apparent motions (Figure 2.1A) The second grouping by UC condition was generated using vertical connecting bars between the simultaneous flashes of each apparent motion and called *vertical connecting bar condition*. These gray vertical connecting bars created a perception of one bigger apparent motion since the simultaneous flashes of each apparent motion was connected with vertical connecting bar and perceived as one single flash (Figure 2.2A). The third grouping by UC condition was introduced as a control condition. It consisted of two vertically aligned apparent motions without any connecting bar between the flashes (Figure 2.3A). This condition is referred as *2 aligned apparent motions condition*. Apparent motions grouped with UC conditions moved to left to right or right to left directions in random order.

Spatial grouping by common fate principle was applied by setting different movement directions for the simultaneous apparent motions. Apparent motion placed upper in the visual field was in the up-down movement direction while the simultaneous other apparent motion (lower one) was in the left-right direction. This condition is referred as *common fate condition* (Figure 2.4A). These two apparent motions had the same parameters as the 2 aligned apparent motions except the direction of motion. Therefore 2 aligned apparent motions condition was used as a control for the common fate condition as well.

Grouping by spatial proximity principle was applied by increasing the vertical distance between two simultaneously presented apparent motions. For the horizontal connecting bar, vertical connecting bar, 2 aligned apparent motions and common fate conditions, two vertically placed simultaneous apparent motions were placed in three different vertical distances: $1.5°$(smallest), $2.7°$(middle) and $4.4°$(largest). Effects of these three proximity conditions on audiovisual interaction were compared for each connecting bar condition (Figures 2.1B, 2.2B, 2.3B, 2.4B).

Each flash of apparent motion was presented for 50 ms. Inter-stimulus interval

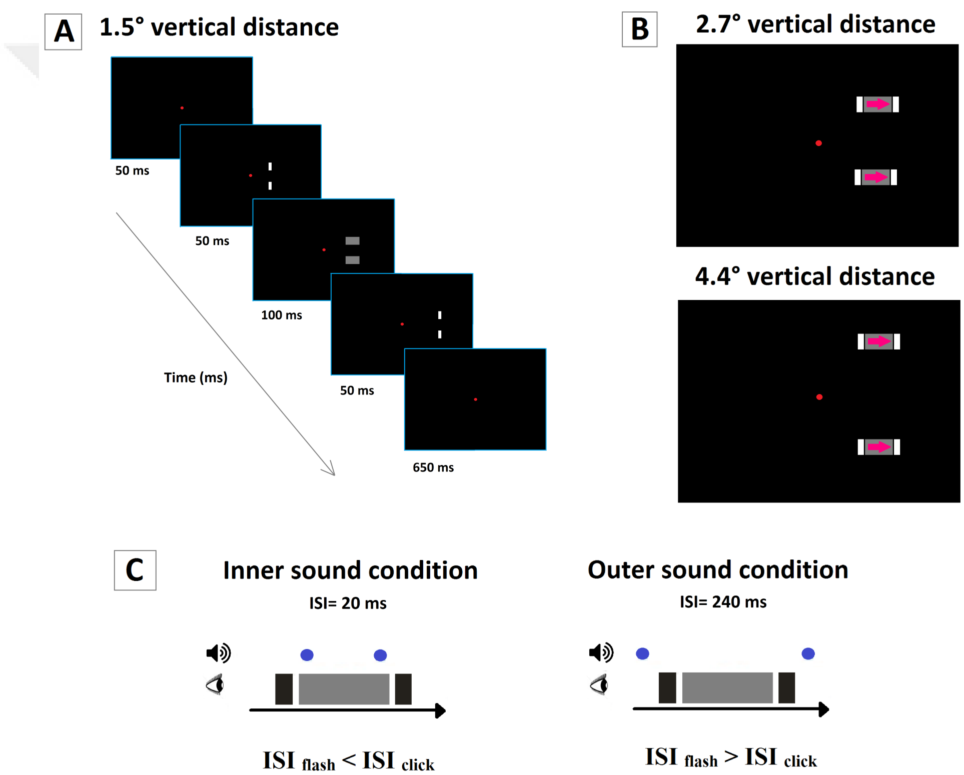

Figure 2.1: Two aligned apparent motions grouped with horizontal connecting bar condition (A). Spatial proximity principle was applied with three different vertical distance parameters (B) Timelines of inner and outer sound conditions relative to the visual stimuli (C)

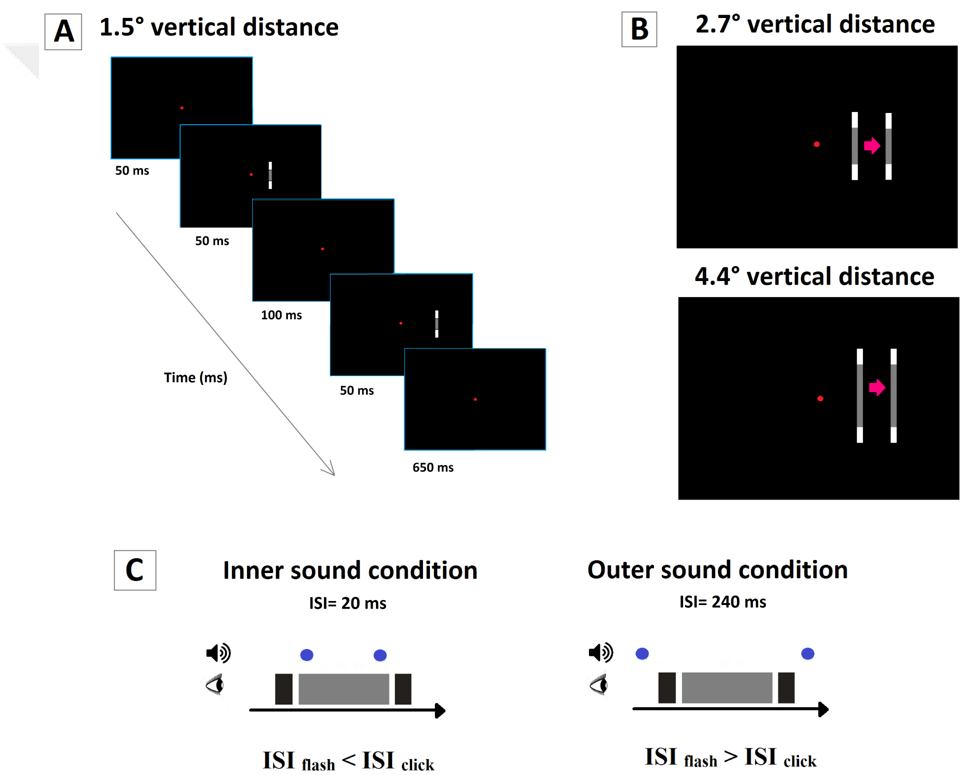

Figure 2.2: Two aligned apparent motions grouped with vertical connecting bar condition (A). Spatial proximity principle was applied with three different vertical distance parameters (B) Timelines of inner and outer sound conditions relative to the visual stimuli (C)

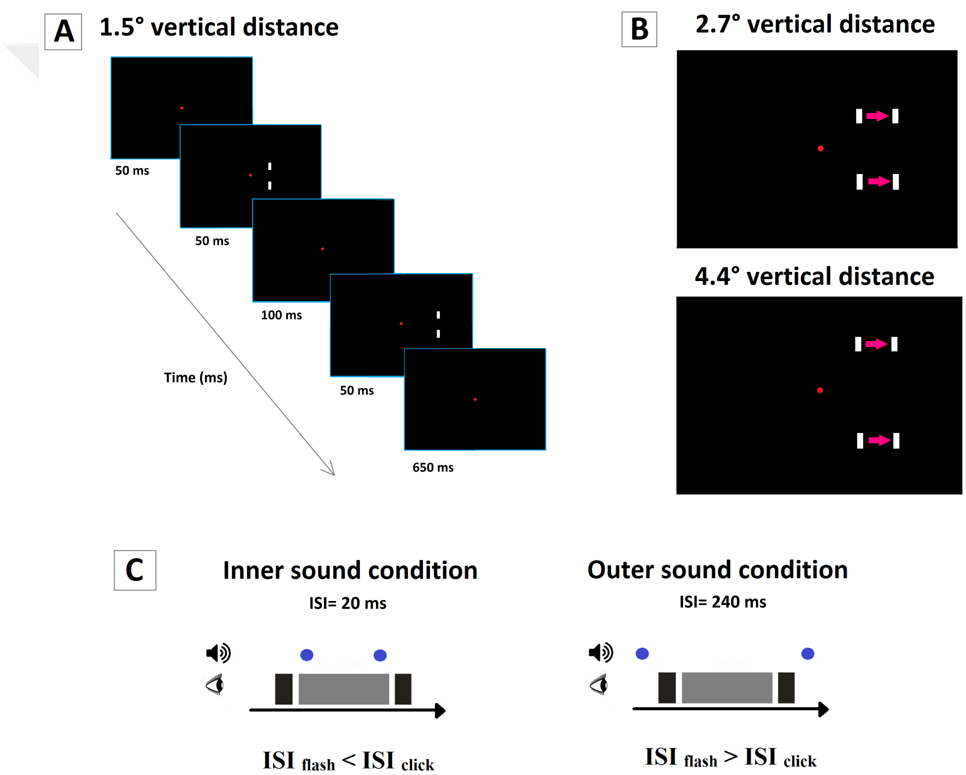

Figure 2.3: Two aligned apparent motions grouped with horizontal connecting bar condition (A). Proximity principle was applied with three different vertical distance parameters (B) Timelines of inner and outer sound conditions relative to the visual stimuli (C)

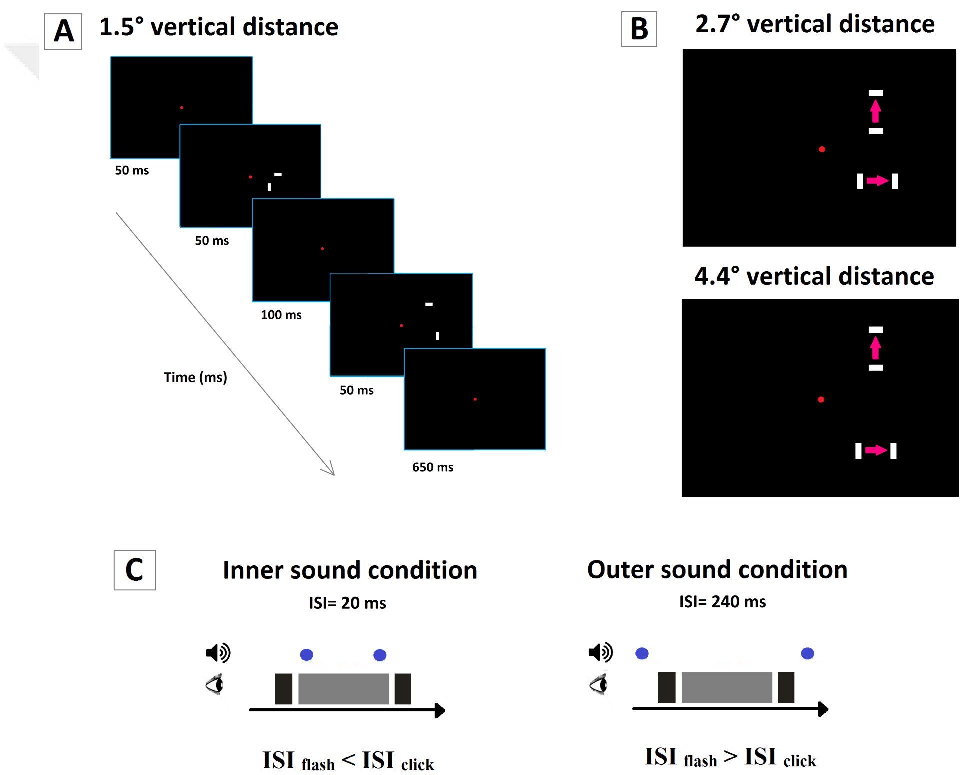

Figure 2.4: Two aligned apparent motions grouped with common fate principle (A). Spatial proximity principle was applied with three different vertical distance parameters (B) Timelines of inner and outer sound conditions relative to the visual stimuli (C)

between the two flashes was 100 ms for all visual grouping conditions. Auditory stimuli were 50 ms clicks, comprised of a rectangular windowed 480 Hz sinewave carrier, sampled at 44.1 kHz with 8-bit quantization. Inner sound condition was composed of 2 clicks per trial with auditory ISI of 20 ms. Outer sound condition, on the other hand, was composed of 2 clicks per trial with auditory ISI of 240 ms. Inner and outer sound conditions were randomly assigned to the first and second sequential apparent motions. The participants were required to listen to these sound conditions delivered through headphones, but make their decision depending on the visual stimuli only. In other words, the participants were informed about the inconsistency between the sound and visual conditions which may mislead their speed judgment.

Multiple apparent motions grouped with UC, proximity or common fate principle were presented twice. These sequentially presented multiple apparent motions were selected from the same visual grouping condition. In other words, if the preceding multiple apparent motions were from the horizontal connecting bar condition, following multiple apparent motions were also from the horizontal connecting bar condition. The participants were instructed to remain their gaze on the fixation point while paying attention to these apparent motions. They were asked to make speed judgment between these sequentially presented clusters of apparent motions by simply reporting the faster one. Responses on this 2 interval forced choice task were collected via a standard keyboard. Hitting the '4'key, reported that first apparent motion was faster than the second, and hitting the '6'key, reported that following apparent motions was faster than the preceding ones. After the response, another set of multiple apparent motions, that were selected randomly from either visual grouping condition, were presented.

Prior to the experimental session, all participants had to take a training session first. The training session consisted of all visual grouping conditions. The sound conditions were not included in the training phase. Basically, the participants were asked to indicate the number of apparent motion they have perceived for random presentation of all visual grouping conditions. The participants reported their decision through the key '1'(reported one apparent motion perceived) and key '2'(reported two apparent motions perceived) on the standard keyboard.

With the training phase, participants familiarity to the visual condition was ensured. Participants who failed to perceive the apparent motion illusion were excluded from the study after the training phase.

Data Analysis

The difference in percentage of inner sound condition seen as faster on vertical connecting bars, horizontal connecting bars, 2 aligned apparent motions and common fate conditions were measured in different spatial proximity parameters. Effects of connecting bar conditions (with 3 levels: vertical, horizontal and no-connecting bars) and spatial proximity conditions (1.5°, 2.7°and 4.4°spatial aperture) on audiovisual interaction were measured with two-way repeated measures ANOVA test. Further statistical analyses assessing the comparison between connecting bar groups for each proximity parameter were conducted with separate two tailed paired-samples t-tests. In order to measure the effect of common fate principle in different proximity conditions, another two-way repeated measures ANOVA was conducted on common fate and spatial proximity conditions. Individual t-tests indicated the difference between common fate conditions on each spatial proximity condition. Note that two separate ANOVA test were needed since connecting bars condition were not compared to the common fate conditions.

Results

Two-way repeated measures ANOVA revealed a significant main effect of connecting bars ($F(2,26)=17.300$, $p<.000$, $\eta_p^2=.57$, $\varepsilon=.665$) suggesting the difference in the effect of sound on apparent motions with horizontal connecting bars, vertical connecting bars and 2 aligned apparent motions. Main effect of spatial proximity conditions s, on the other hand, was not found as statistically significant, $p>.05$. Audiovisual interaction did not seem to change across different spatial proximities where each apparent motion was placed. However, the interaction between the spatial proximity and the connecting bar conditions was found as statistically significant ($F(4,52)=3.429$, $p=.015$, $\eta_p^2=.21$, $\varepsilon=.623$). The effect of connecting bars on audiovisual interaction depended on the levels of the spatial proximity between the apparent motions. An interaction effect is evident in Figure 2.5, where simple main effect of connecting bars changed over different spatial proximity conditions, specifically between vertical connecting bars and 2 aligned apparent motions conditions on 2.7°and 4.4°. Moreover, the difference between the horizontal and vertical connecting bar conditions seemed to differ at a great deal particularly on 2.7°and 4.4°of proximity conditions compared to 1.5°. Means and standard errors of connecting bar and proximity conditions are demonstrated in 1 (Appendix B).

Further analyses regarding the comparison between levels of spatial proximity and connecting bar conditions were conducted with individual paired-samples t-tests. In smallest spatial proximity condition (1.5°) effect of sound was significantly larger for 2 aligned apparent motions condition relative to the horizontal connecting bar condition ($t(13)$=-2.759, p=.016) (Bonferroni corrected). The difference in percentage of inner sound condition seen as faster was not significantly different between horizontal and vertical connecting bar conditions ($t(13)$=-1.768, $p>$.101) and between vertical connecting bar and 2 aligned apparent motions conditions($t(13)$=-.551, $p>$.591). In spatial proximity of 2.7°, significantly smaller percentage of inner sound condition seen as faster was observed for horizontal connecting bar condition relative to the 2 aligned apparent motions condition ($t(13)$=6.558, p=.000) and vertical connecting bar condition ($t(13)$=-4.990, p=.000) (Bonferroni corrected). The vertical connecting bar and 2 aligned apparent motions conditions were not significantly different ($t(13)$=-1.087, p=.297). At the largest vertical distance (4.4°), significantly smaller percentage of inner sound condition seen as faster was found for horizontal connecting bar condition relative to vertical connecting bar condition ($t(13)$=3.606, p=.003)(Bonferroni corrected). The difference between vertical connecting bar and 2 aligned apparent motions conditions at the spatial proximity of 4.4°was not significant ($t(13)$=2.161, p=.05)(Bonferroni corrected).

In the spatial proximity of 2.7°, the differences in percentage of inner sound perceived as faster between the connecting bar conditions were consistent with our hypothesis. As expected, the horizontal connecting bar condition was associated with less auditory capture compared to the vertical connecting bar and

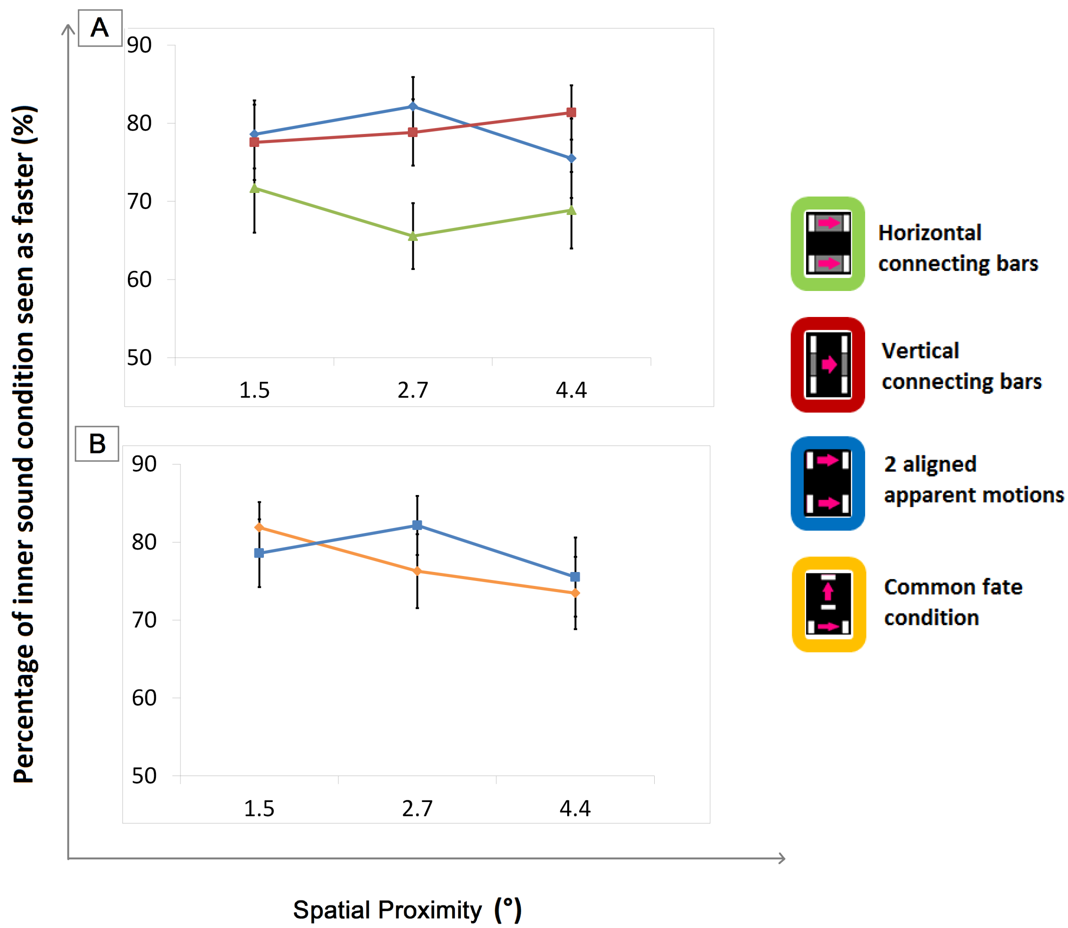

Figure 2.5: Difference in percentage of inner sound condition seen as faster on two aligned apparent motions grouped with different uniform connectedness principles (A) and common fate principles (B) for three levels of proximity condition. Error bars indicate ± SEM.

2 aligned apparent motions conditions. It indicated that vertically grouped apparent motions were most likely perceived as one; therefore they were captured by the accompanying sounds more than the horizontal connecting bar condition which were most likely perceived as two distinct apparent motions and influenced from the clicks significantly less. As expected, vertical connecting bar and 2 aligned apparent motions conditions was not significantly different in terms of auditory capture. Figure 2.5 also demonstrates that for all spatial proximity parameters, the percentage of the inner sound condition seen as faster was lower in the horizontal connecting bar condition relative to the 2 aligned apparent motions and vertical connecting bar conditions. Thus, significantly less effect of auditory capture in the horizontal connecting bar condition was present for all spatial proximity parameters.

In order to investigate the effect of spatial grouping by common fate and proximity principles on audiovisual interaction, two vertically aligned apparent motions moving in left-right direction was compared to the two vertically aligned apparent motions moving to different directions for each spatial proximity parameter. Two-way repeated subjects ANOVA did not indicate significant main effect of common fate condition (p>.05). However main effect of spatial proximity was found as statistically significant, F(2,26)=4.219, p=.03, η_p^2=.24, ε=.860. The interaction effect between the common fate and spatial proximity conditions was also not statistically significant (p>.05). Means and standard errors of common fate and proximity conditions are demonstrated in Table2 (Appendix B). Individual paired-samples t-tests were conducted between the common fate and 2 aligned apparent motions conditions for each spatial proximity condition. Common fate and 2 aligned apparent motion conditions were not found statistically significant from each other at any spatial proximity condition, p>.02 (Bonferroni corrected).

Figure 2.5B indicates the differences in percentage of inner sound condition seen as faster between the 2 aligned apparent motions and common fate conditions on each spatial proximity conditions. In accordance with our predictions, a trend of decreasing auditory effect with increasing spatial proximity was observed in the common fate condition. In fact, percentage of inner sound condition seen as faster was found to be larger in 4.4°of spatial proximity relative to 1.5°for common fate conditions ($t(13)=2.989$, $p=.010$) (Bonferroni corrected). Nonetheless, blue line indicating 2 aligned apparent motions condition and orange line indicating common fate condition did not reveal a consistent trend across vertical distances.

Overall, experiment 1 indicated that visual stimuli grouped with spatial proximity and common fate principles did not have a consistent effect on the audiovisual interaction. Only the spatial organization of uniform connectedness principle (horizontal and vertical connecting bars) successfully facilitated or decrease the effect of sound over the visual stimuli. Experiment 2 was designed in order to generalize the effect of uniform connectedness observed in experiment 1 to increased number of visual stimuli. Furthermore, experiment 1 measured the effect of sound on sequential apparent motions with equal ISIs between the flashes. Experiment 2 further address the effect of sound over apparent motions with different ISI values. In other words, it is investigated whether physically faster apparent motions would be perceived as slower when outer sound clicks accompany. Details of experiment 2 were explained below.

2.2 Experiment 2- Three Apparent motions grouped with Uniform Connectedness Principle

This experiment was designed to generalize the effect of spatial organization with uniform connectedness on audiovisual interaction to more than two visual stimuli. Horizontal and vertical connecting bars were used to generate two different spatial organizations which either facilitate or decrease the effect of sound conditions. Sequential apparent motions were varied in ISI in order to investigate the effect of sound conditions over physically faster and slower apparent motions.

Subjects

Subjects were 6 volunteers (3 female 3 males) naïve to the purpose of the experiment. All participants had intact or corrected to normal visual acuity and intact hearing. Participants gave informed consent of participation and all procedures were in accordance with international standards (WMA Declaration of Helsinki) and approved by Ankara University Ethics Committee.

Stimuli and Procedure

The apparatus, auditory stimuli and screen parameters were the same as the ones used in the first experiment. In the second experiment, the visual stimuli consisted of three vertically placed, sequential flashes (0.3 x 1.2 degree, with a luminance of 35 cd/m^2) with fainted gray bars (with a luminance of 35 cd/m^2) generating three simultaneous apparent motions. These simultaneous apparent motions moved in left-right direction and they were presented on a black background. Three simultaneous apparent motion stimuli with horizontal connecting bars between the sequential flashes formed the first grouping condition called *horizontal connecting bar condition* (Figure 2.6A). The second grouping condition was generated with vertical connecting bar between the simultaneous first and second flashes of each apparent motion and called *vertical connecting bar condition* (Figure 2.6B). The horizontal and vertical fainted gray connecting bars created two spatial organization conditions. While the gray horizontal connecting bars created a perception of three distinct simultanous apparent motions; the gray vertical connecting bars created a perception of one bigger apparent motion.

The spatial proximity and common fate conditions employed in the first experiment were not used in the current experimental design since they did not have a significant effect on audiovisual interaction. Instead, spatial proximity parameter between the three vertically aligned apparent motions was set to 2.7°.

Auditory stimuli were 20 ms clicks, comprised of a rectangular windowed 480 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization. Inner sound condition was composed of 2 clicks per trial with auditory ISI of 20 ms and

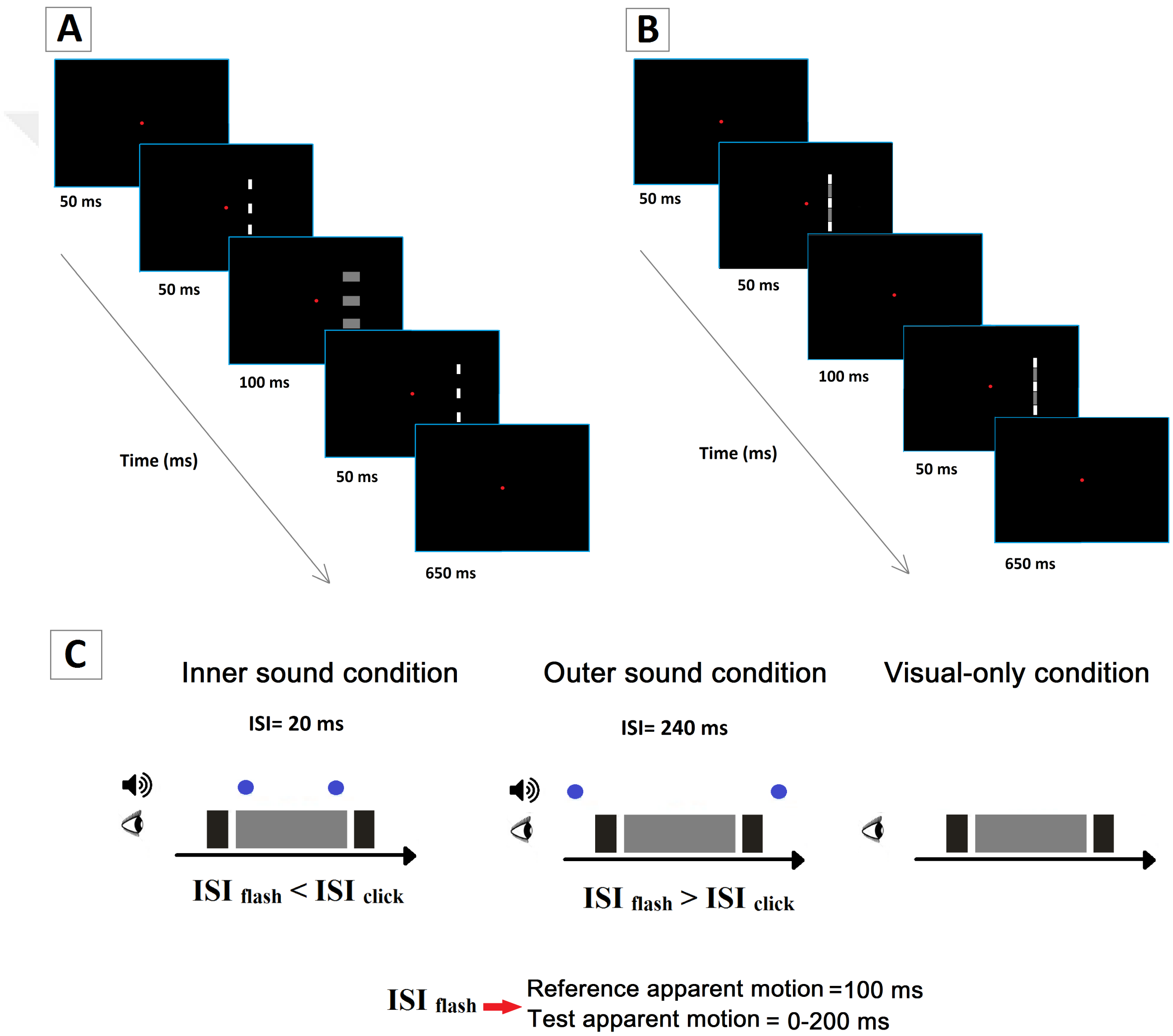

Figure 2.6: Three apparent motions grouped with horizontal (A) and vertical connecting bars (B). Timelines of the inner, outer and visual-only conditions (C)

expected to increase the perceived speed of apparent motion. Outer sound condition, on the other hand, was composed of 2 clicks per trial with auditory ISI of 240 ms and expected to decrease the perceived speed of apparent motion.

Procedure was similar to the first experiment. 2 interval forced choice task consisted of sequentially presented clusters of three simultanous apparent motions which were presented on the right hemi-field of the display. The first three had an ISI value fixed to 100 ms between the flashes. These particular apparent motions were referred as the *reference apparent motions*. The second three apparent motions had a variable ISI value of 0-200 ms between the flashes and these apparent motions were referred as *test*. Test apparent motions with 0 ms ISI interval created a perception of the fastest apparent motion (simultaneous flashes) while 200 ms ISI interval created the slowest. Test and reference apparent motions were randomly presented, meaning that preceding apparent motion could be either test or reference. Inner and outer sound conditions only accompanied to the test apparent motions.

All participants had completed the training session prior to the experimental session. The training session was the same as the one used in the first experiment.

Data Analysis

For each visual grouping condition, group averaged data was fitted by a Psychometric Curve, using psignifit function in MATLAB [43]. The 50% point on sound based responses represented the *Point of Subjective Equality (PSE)* in which participants could not discriminate which clusters of apparent motions were faster. For each sound conditions, ISI intervals corresponding to PSE points were calculated. PSE for visual-only condition corresponds to the ISI of 100 ms which is the interval at which speed judgments were reported equally probable. Sound induced changes in PSE values were compared separately for horizontal and vertical connecting bar conditions. The difference between PSE values of the inner and outer sound conditions were defined as temporal ventriloquism effect (TVE).

Results

The effect of uniform connectedness on audiovisual interaction was analyzed with two-way repeated-measures ANOVA. Main effect of sound on perceived speed was found statistically significant ($F(1,5) = 545.596$, p<.000, η_p^2=.99, ε=1) and the interaction between the sound and connecting bar conditions was found statistically significant ($F(1,5) = 17.821$, p= .008, η_p^2=.78, ε=1). However the main effect of connecting bar condition was not found statistically significant (p<.05). Means and standard errors of connecting bar and sound conditions are demonstrated in Table3 (Appendix B).

Group averaged data for the horizontal connecting bar condition indicated less shift of ISI value between the inner and outer sound conditons (Figure 2.7A) compared to the vertical connecting bar condition (Figure 2.7B). When the inner and outer sound conditions accompanied the three apparent motions with horizontal and vertical connecting bars, the PSE values were significantly different across the connecting bar conditions. Individual paired-samples t tests indicated that the PSE values of the inner sound conditions was significantly smaller than the outer sound condition for the horizontal connecting bar ($t(5)$=-9,638, p<.000) and the vertical connecting bar conditions ($t(5)$=-19,383, p<.000) (Bonferroni corrected) (Figure 2.8A). Moreover the TVE value in the horizontal connecting bar condition was found significantly smaller from the TVE value in vertical connecting bar condition ($t(5)$=-4.221, p=.008)(Bonferroni corrected). Figure 2.8B indicates larger TVE value found in the vertical connecting bar compared to the horizontal connecting bar condition. In addition, the inner sound conditions were found significantly larger in horizontal connecting bar condition relative to the vertical connecting bar condition ($t(5)$=3.782, p=.013)(Bonferroni corrected). However, the outer sound conditions were not significantly different across the horizontal and vertical connecting bar conditions (p>.02).

Overall, experiment 2 indicated that horizontal connecting bars decreased the effect of auditory capture while vertical connecting bars facilitated it. Greater shift of PSE value observed in vertical connecting bar condition demonstrated that objects grouped as one are more likely to be influenced from the inner and

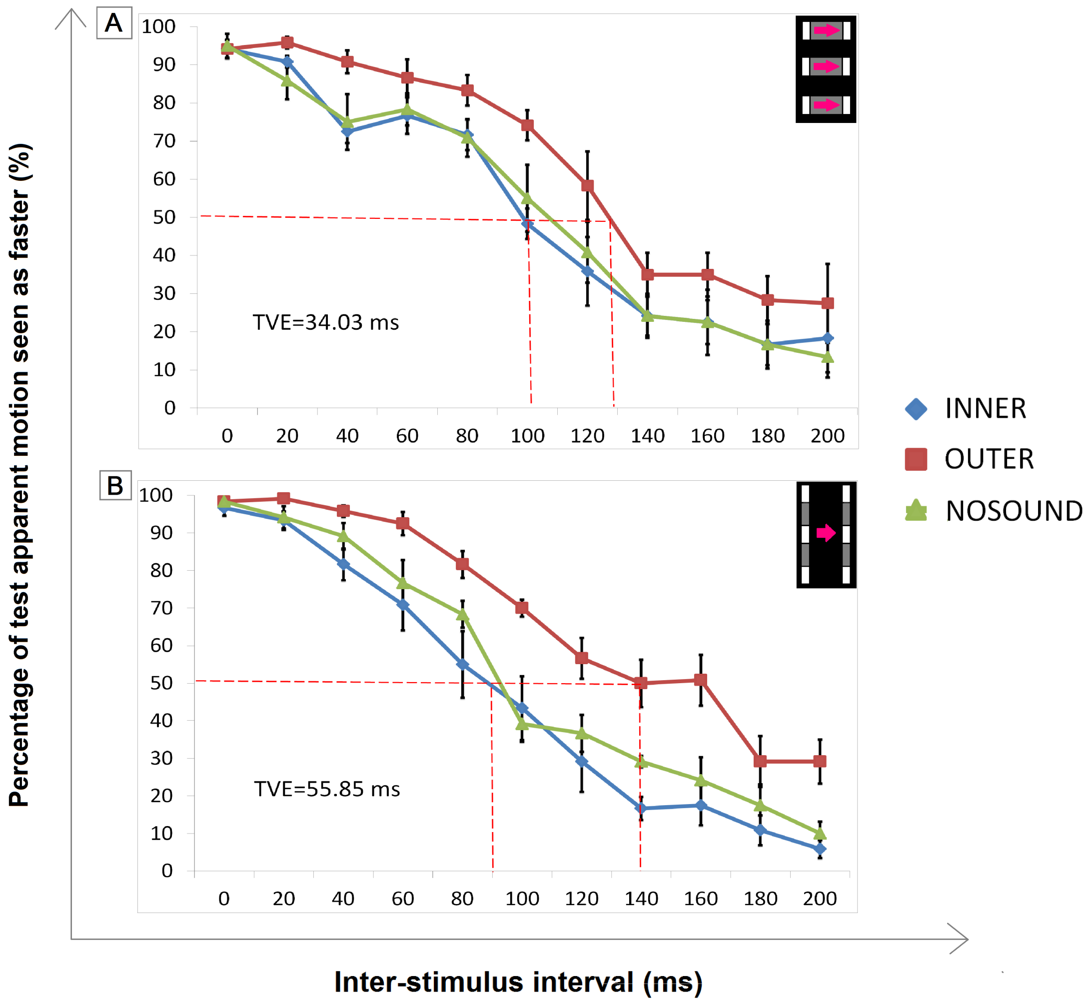

Figure 2.7: Group averaged raw data in three apparent motion clustered with horizontal connecting bars (A) and vertical connecting bars (B). Error bars indicate ± SEM.

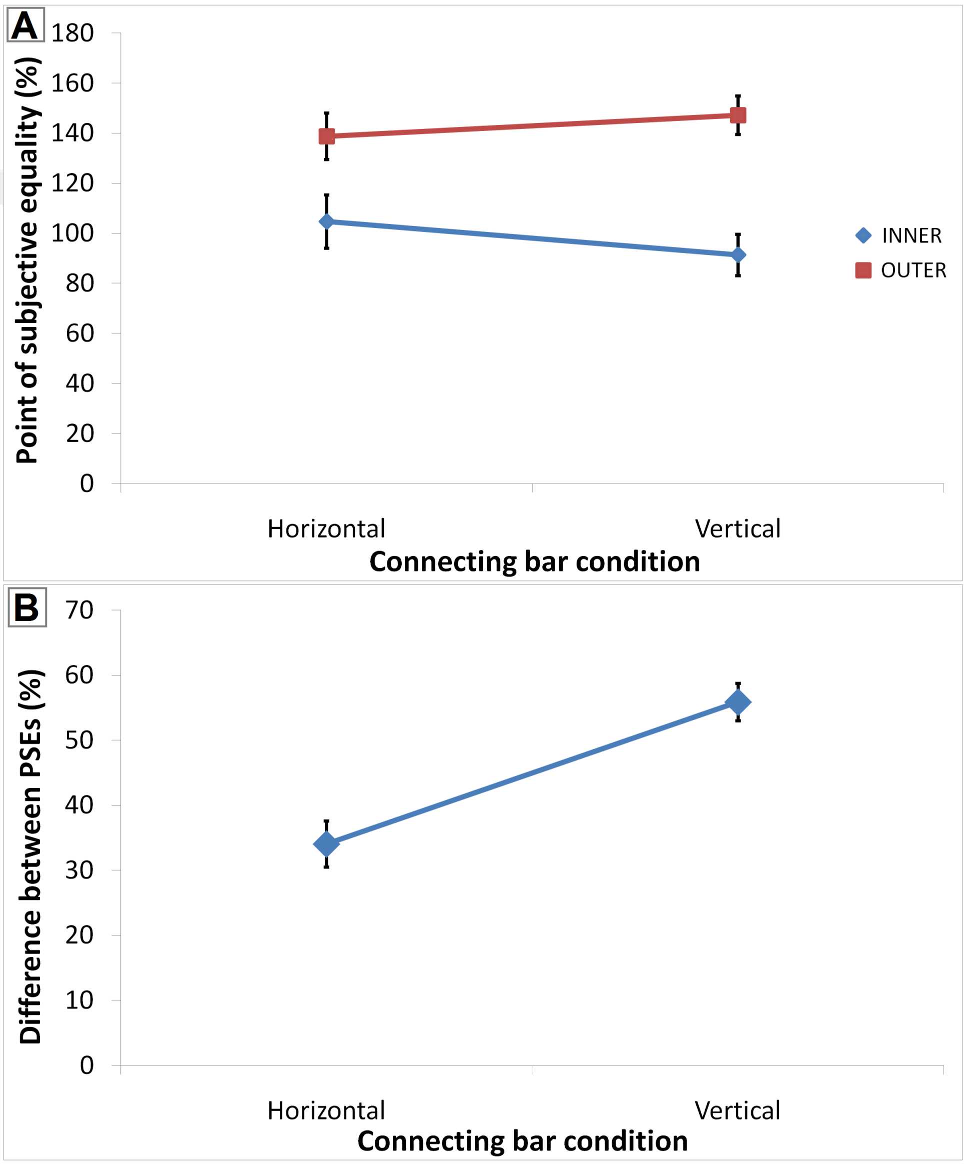

Figure 2.8: PSEs of sound conditions (A) and TVE scores (B) across horizontal and vertical connecting bar conditions. Error bars indicate ± SEM.

outer sound conditions. In vertical connecting bar condition, physically slower apparent motions were perceived faster with inner sound condition and physically faster apparent motions were perceived slower with outer sound condition. Similar shift in perceived speed was present but significantly smaller in horizontal connecting bar condition. Connecting bars modulated the audiovisual interaction with multiple visual stimuli at a great deal but whether these results could be generalized to single visual stimulus was unknown. Thus, experiment 3 applied these connecting bars to single apparent motion stimuli and investigated the change in auditory capture.

2.3 Experiment 3- Single Apparent Motion Grouped with Uniform Connectedness Principle

Experiment 3 was conducted to investigate whether the less effect of the sound conditions over the apparent motions grouped with horizontal connecting bars observed in the first two experiments could be generalized to the single apparent motion stimulus as well. It was investigated whether physically slower apparent motions would be perceived as faster with inner sound condition while physically faster apparent motions would be perceived as slower with outer sound conditions when these apparent motions were grouped with horizontal connecting bar. Note that in experiment 3, the horizontal connecting bar condition was compared to the baseline condition (no-connecting bar) rather than the vertical connecting bar condition.

Subjects

Participants were 8 volunteers (5 female, 3 male) from Bilkent University. All participants had intact or corrected to normal visual acuity and intact hearing. Participants gave informed consent of participation. All procedures were in accordance with the international standards (WMA Declaration of Helsinki) and approved by the ethics committee of Ankara University.

Stimuli and Procedure

Auditory stimuli parameters and the apparatus were similar to the one in the second experiment. However, instead of multiple vertically aligned apparent motions, only one apparent motion moving in left-right direction was used in this experiment. Apparent motion stimulus with a horizontal connecting bar between the flashes formed the first grouping condition and referred as *single horizontal connecting bar condition* (Figure 2.9A). The second grouping condition was created as a control which had the same apparent motion but without the horizontal connecting bar. This condition was referred as *single baseline condition*(Figure 2.9B).

The procedure was also similar to the one in the second experiment. In the 2 interval forced choice task, the test apparent motion (ISI between flashes varied over 0-200 ms) and the reference apparent motion (ISI between flashes was fixed to 100 ms) were introduced in a random order. The auditory stimuli only accompanied the test apparent motion.

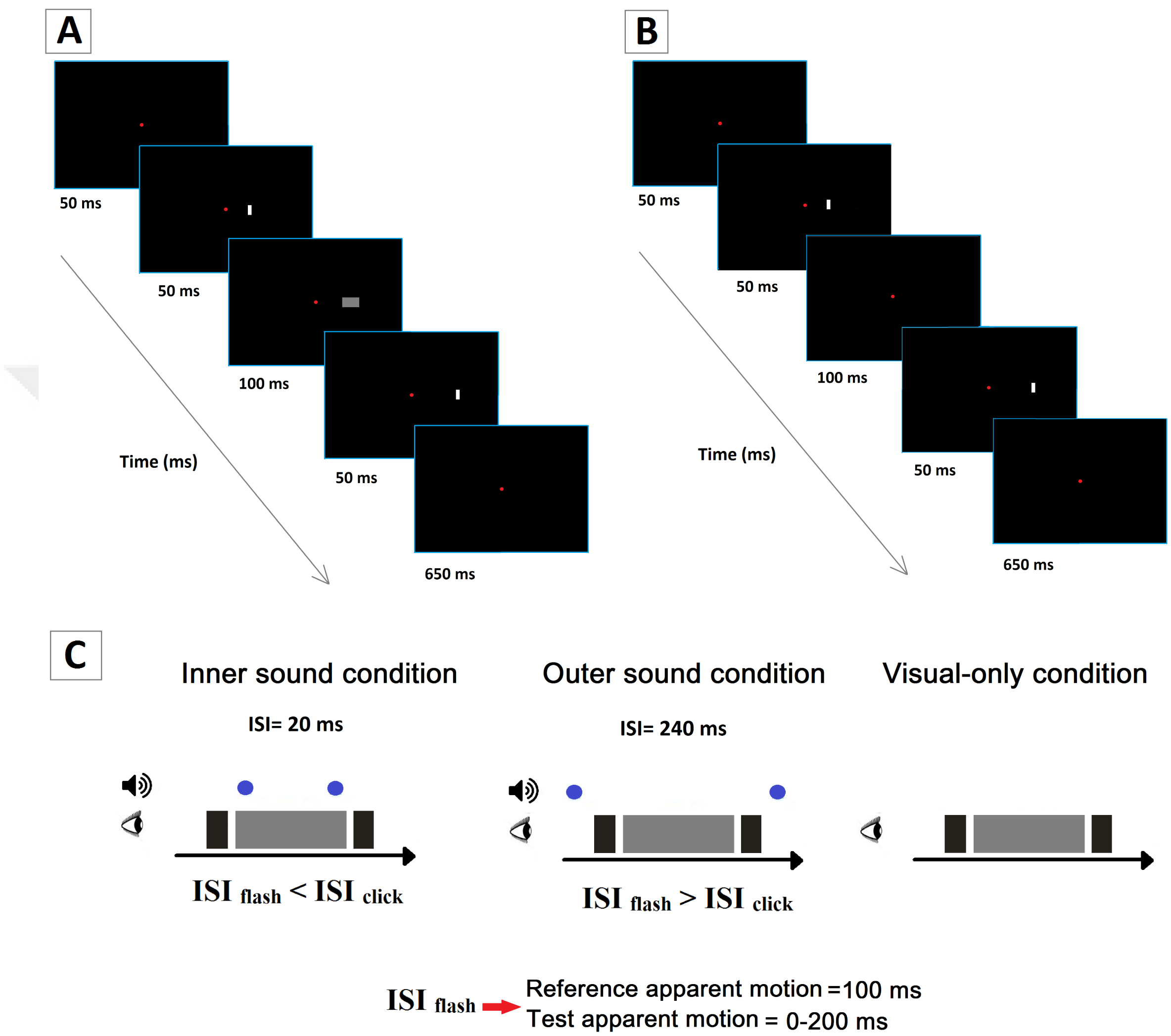

Figure 2.9: Single horizontal connecting bar (A) and single baseline conditions (B). Timelines of inner, outer and no-sound conditions (C)

Data Analysis

Psychometric curves of the group averaged data were fitted for each sound condition and visual grouping condition. The TVEs were calculated by subtracting the PSE values of the inner and outer conditions and they were compared across the connecting bar conditions. The PSE for the visual-only condition was calculated as ISI of 100 ms. The shift in the ISI values due to the inner and outer conditions were compared to ISI value of 100 ms (visual-only condition).

Two-way repeated measures ANOVA and paired-samples t-tests were performed using SPSS.

Results

Two-way repeated measures ANOVA indicated statistically significant main effect of the sound condition ($F(1{,}7)$=65.777, p=.005, η_p^2=.70, ε=1) and statistically significant interaction effect of connecting bar and sound conditions ($F(1{,}7)$=6.824, p=.04, η_p^2=.47, ε=1). However the main effect of connecting bar condition was not statistically significant ($p>.05$). Means and standard errors of connecting bar and sound conditions are demonstrated in Table4 (Appendix B).

Figures 2.10A and 2.10B indicate the shift in the ISI values corresponding to the 50% test apparent motion seen as faster responses for the inner and outer sound conditions with respect to visual-only condition in the single apparent motion with and without connecting bars. At the point where the sequential apparent motions were judged to be equally fast, outer sound condition indicated a larger ISI value which led to a percept of slower apparent motion and inner sound indicated a smaller ISI value which led to a percept of faster apparent motion. Similar shift of ISI value relative to the visual-only condition was observed not only on the PSE value, but also around physically smaller and larger ISIs. The outer sound condition pulled the smaller ISI values towards the larger ones while the inner sound condition pulled the larger ISI values to the smaller ones, compared to the visual-only condition. These shifts in ISI values due to the sound conditions were found to be smaller for the single horizontal connecting bar condition compared to the baseline condition.

Individual paired-samples t-tests indicated that the outer sound had significantly larger PSE value compared to the inner sound conditions for both single horizontal connecting bar ($t(7)$=-2.859, p=.024) and single baseline conditions ($t(7)$=-4.149, p=.004) (Bonferroni corrected). It was demonstrated that the inner-outer difference for the single baseline condition was larger compared to the single horizontal connecting bar condition (Figures 2.10A, 2.11A). The inner

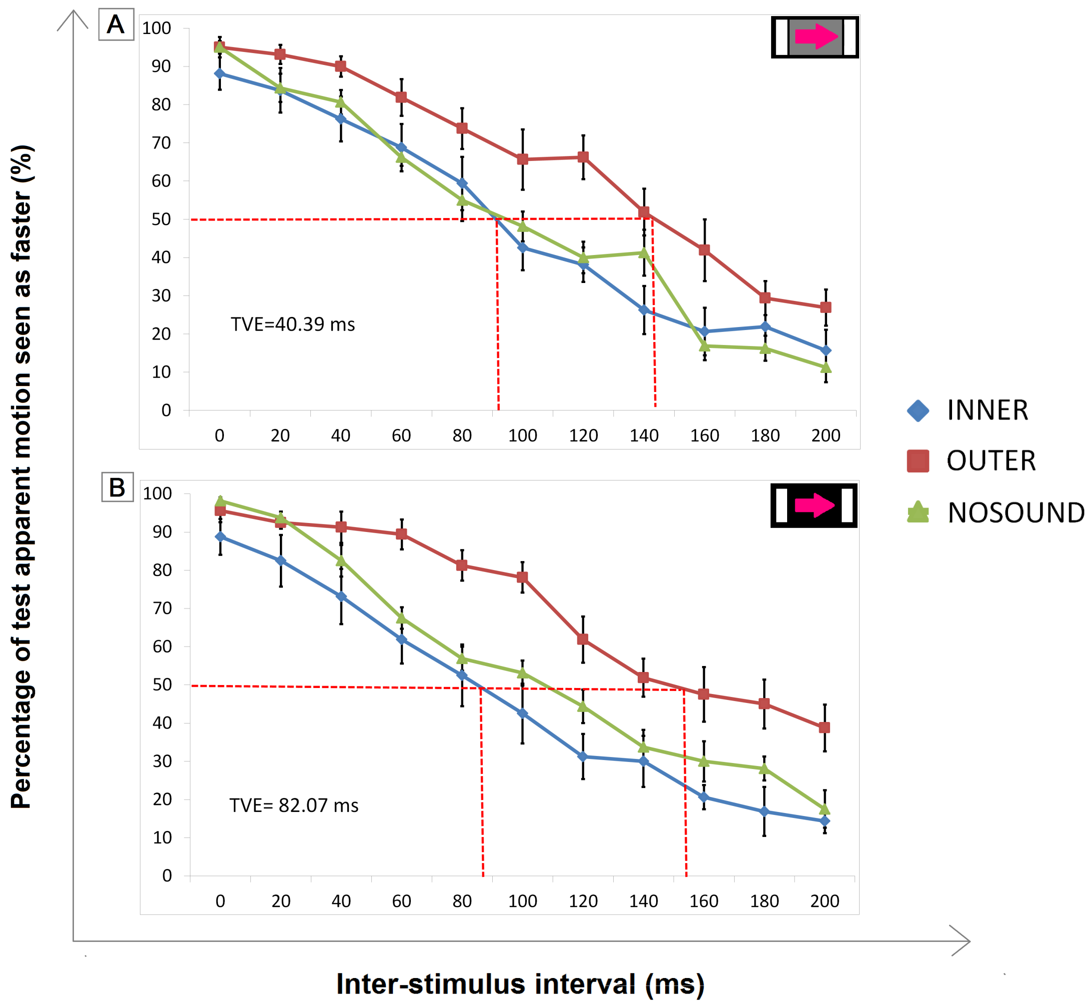

Figure 2.10: Group averaged raw data in single apparent motion with horizontal connecting bar (A) and baseline conditions (B). Error bars indicate ± SEM.

sound condition was not significantly different across the connecting bar conditions (t(7)=1.388, p=.208). Also the outer sound condition was not significantly different across connecting bar conditions (t(7)=-2.599, p=.035) (Bonferroni corrected).

Furthermore, paired-samples t-test indicated that significantly larger TVE score in baseline condition relative to the single horizontal connecting bar condition (t(7)=-2,510, p=.040) (Figure 2.11B). It was observed that in the presence of connecting bars, the participants grouped the sequential flashes of the apparent motion and they perceived the whole event as one. Therefore, two static clicks were unable to capture the final continuous apparent motion. On the other hand, when the horizontal connecting bar was not present, two sequential flashes of apparent motion bound to two static sounds. Therefore, for the single baseline condition, the speed judgment depended more on the accompanying static sounds.

Overall, experiment 3 indicated that horizontal connecting bar introduced between the flashes of single apparent motion created a continuous apparent motion illusion which was perceived as a single object moving back and forth. Since each sound condition consisted of two static clicks and visual stimulus consisted of a single object, effectiveness of inter-modal binding decreased. In case of baseline condition, two flashes of apparent motion successfully binded to the clicks of auditory conditions and facilitated the audiovisual interaction. Therefore when outer sound condition accompanied to the baseline condition, physically faster apparent motions were perceived as slower and when inner sound accompanied, physically slower apparent motions were perceived as faster relative to the visual-only condition.

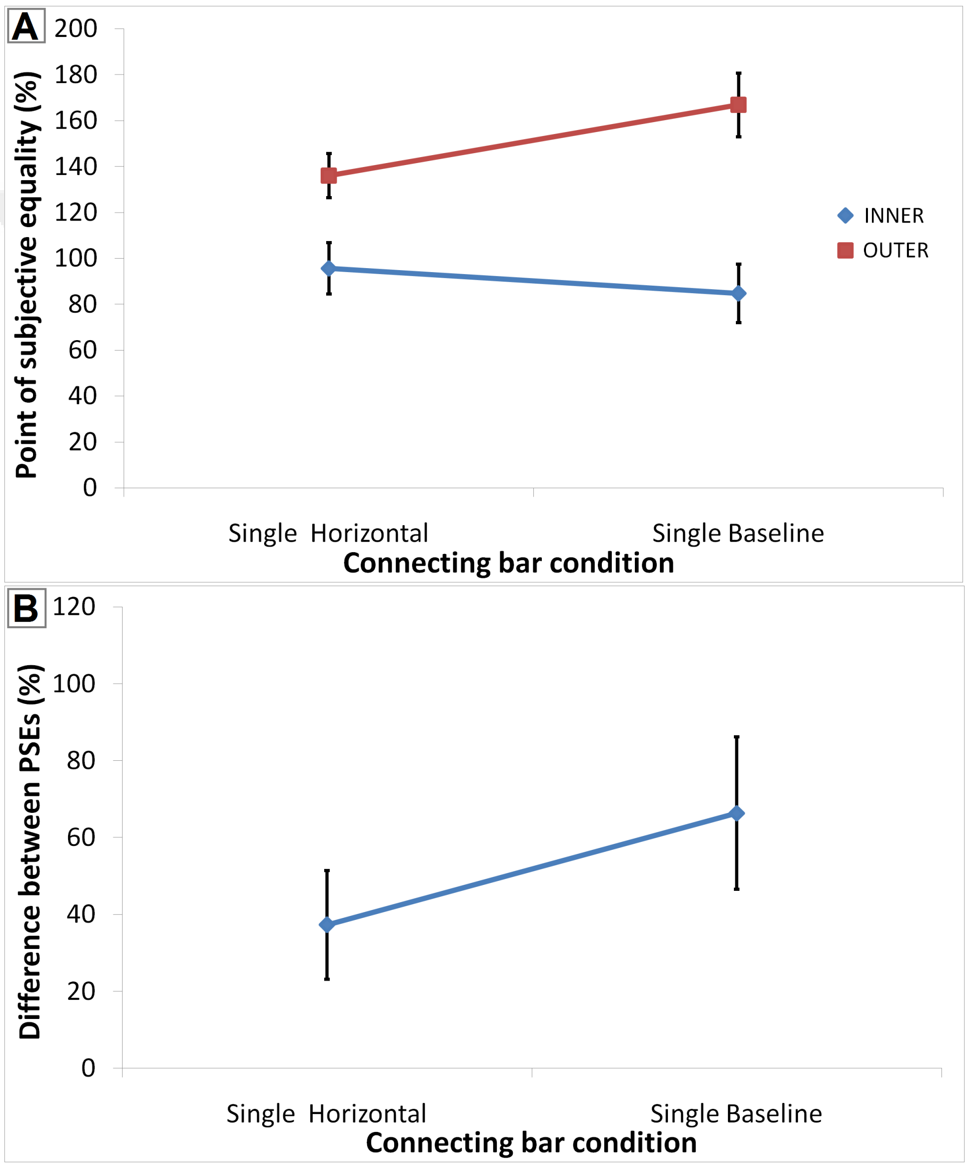

Figure 2.11: PSEs of sound conditions (A) and TVE scores (B) compared across single apparent motion with horizontal connecting bar and baseline conditions. Error bars indicate ± SEM.

Chapter 3

EEG Experiments: Method and Result

Behavioral experiments indicated auditory capture over visual stimuli grouped with uniform connectedness principle. Aim of the EEG experiment was to measure possible differences in amplitude of averaged ERPs from apparent motion/s grouped with horizontal and vertical connecting bars. Uni-modal conditions (visual only and auditory only) were compared to audiovisual conditions on both sensory specific and non-specific areas.

3.1 EEG Recording and ERP Analysis

Subjects

Participants were 12 (8 male 4 female) adults with intact or corrected vision and normal hearing. All gave a written informed consent before participating to the experiment. EEG data was acquired from each participant at one session only. All participants went through training with the prior behavioral data acquisition and all were familiar with the task.

Stimuli

Stimulus presentation was implemented with Matlab 2012b, under same settings as behavioral data acquisition. Continuous EEG data was acquired in a dim lighted room in order to eliminate noise induced by excessive alpha waves. Raw EEG data consisted of audio-visual, visual-only and auditory-only conditions. On audiovisual and visual-only conditions, two aligned apparent motions with horizontal, vertical connecting bars and without connecting bars (2 aligned apparent motion) and single apparent motion with and without horizontal connecting bar were presented. Two sound conditions (inner and outer sound) randomly accompanied sequential apparent motion/s during audiovisual trials. Auditory-only condition included only random presentation of inner and outer sound conditions. Participants were asked to report the faster apparent motion by pressing 4 or 6 on the keyboard as they did in the behavioral experiments. They were asked to rely their speed judgment on the clicks only during auditory-only trials.

Apparatus

EEG data was recorded through 64 channels Easy Cap GmbH and amplified with BrainAmp, Brain Products, Germany. Recording was at a sampling rate of 5000 Hz. Prior to the data acquisition, electrode impedances were maintained below 20 kΩ. Auditory stimuli were presented through E-A-RTONE Gold 3A Insert Earphones.

ERP Analysis

Offline data preprocessing was performed with BrainVision Analyzer 2 (Brain Products, Germany) and Matlab R2016a. Raw EEG data was down-sampled to 500 Hz. Data was offline low-pass filtered at 100 Hz, high pass filtered at 0.5 Hz and notch filtered at 50 Hz. After IIR filters, data was segmented into epochs of 600 ms, from -100 ms to 500 ms. Independent component analysis was carried out in order to remove possible EEG artifacts such as eye blinks. Each trial was examined both by an automatic artifact rejection criteria and eye manually; trials with artifacts were discarded. Trials with oscillations over 50 µV/ms and trials

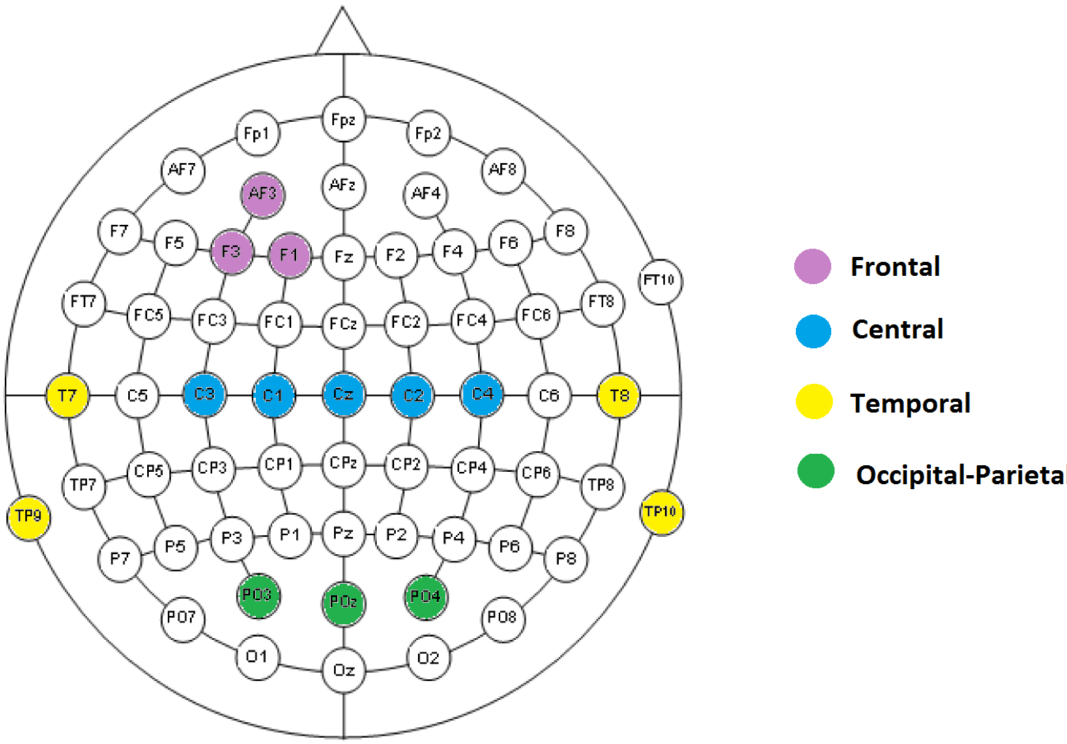

Figure 3.1: ROIs determined by individual two tailed t-tests between all 63 channels

with voltage change over 200 µV or having oscillations 0.5 µV in 100 ms were eliminated by automatic rejection criteria.

The ERPs were time locked to onset of visual stimulus and they were averaged separately for each connecting bar condition (horizontal connecting bar, vertical connecting bar, 2 aligned apparent motion, single horizontal connecting bar, single baseline) and for each sound condition (inner, outer, visual-only). The averaged ERPs obtained in auditory-only, visual-only and audiovisual conditions were exported to Matlab and individual two tailed t-tests on each time points between all 63 channels were conducted to define ROIs (see Appendix A). Four ROIs from frontal, temporal, central and occipital-parietal regions were generated by selecting channels that were significantly activated in response to auditory-only, visual-only and audiovisual conditions (Figure 3.1).

First analysis focused on amplitude change on each time point from 100 ms before stimulus onset to 500 ms post-stimulus onset associated with audiovisual (AV) inner and outer sound conditions on all connecting bar conditions in order to assess whether apparent motion/s accompanied by inner sound would differ from the apparent motion/s accompanied by the outer sound condition in terms of their perceived speed. Individual paired-samples t-tests between AV inner and AV outer were conducted. Next analysis compared amplitude change in audiovisual minus auditory-only conditions to visual-only conditions (AV-A vs Visual-only) on -100 ms to 500 ms time window in order to observe the individual effect of sound conditions on connecting bar conditions. Auditory-only inner sound condition was subtracted from AV inner sound condition and auditory-only outer sound condition was subtracted from AV outer sound condition. Then these two AV-A components were compared to the visual-only conditions with individual paired-samples t-tests.

3.2 EEG Results

For all ROIs, the difference between AV inner and AV outer sound conditions and the effect of individual sound conditions on each connecting bar condition were investigated. Averaged ERPs associated with inner sound, indicated in red line, and outer sound, indicated in blue line, were compared and the time intervals in which both sound conditions were found significantly different from each other were marked with black solid lines (Figures 3.2, 3.4, 3.6, 3.8, A and B). Visual-only condition (demonstrated in green line) was used to compare the effect of sound conditions on connecting bar conditions.

Frontal Channels

For frontal channels, the difference between AV inner and outer sounds accompanying horizontal connecting bar was found over 52-70 ms, 130-176 ms and 218-282 ms (p<.05). 52-70 ms indicated an earlier time interval where the AV outer sound condition induced significantly larger positive amplitude values relative to the positive amplitude values associated with the AV inner sound condition. However, around 130-176 ms P2 component associated with the AV inner sound condition was significantly larger relative to the AV outer sound condition at later time interval. 218-282 ms time interval, which corresponds to the end of stimulus presentation, larger N1 component was observed for the AV inner sound condition compared to the AV outer sound condition (Figure 3.2A, 1^{st} row). The AV-A inner was found significantly different from the visual-only condition during 258-290 ms (p<.05). Nonetheless the AV-A outer was not found significantly different from the visual-only condition (p>.05)(Figure 3.2B, 1^{st} row).

In vertical connecting bar condition, statistically significant AV inner-outer difference was found over 36-88 ms, 140-170 ms during stimulus presentation and 232-258 ms post-stimulus presentation (p<.05). 36-88 ms time interval covered around larger P1 component associated with the AV outer sound condition compared to the AV inner sound condition. Around 140-170 ms, the AV outer sound and visual-only conditions revealed similar amplitude values (around baseline) while the AV inner sound condition indicated a late positive peak (compared to the P1 associated with AV outer sound condition). 232-258 ms time interval corresponded to the end of stimulus presentation and around the negative peak associated with the AV inner sound condition relative to the AV outer sound condition (Figure 3.2A, 2^{nd} row). Around the late negative peak induced by the AV-A inner sound condition was found to be significantly larger from the visual-only condition during 368-390 ms (p<.05); while the ERP associated with the AV-A outer sound condition did not reveal any statistically significant difference from the visual-only condition (p>.05).(Figure 3.2B, 2^{nd} row).

The AV inner-outer sound difference on two aligned apparent motions condition was found statistically significant over 6-26 ms (around stimulus onset), 136-174 ms and 366-374 ms intervals (p<.05). These intervals included the larger amplitude value of the AV inner sound condition around the stimulus onset and around larger P1 component associated with the AV inner sound condition compared to the AV outer sound condition. Around 366-374 ms, amplitude change associated with the AV inner sound condition returned to baseline while AV outer sound condition induced another negative peak (Figure 3.2A, 3^{rd} row). Statistically significant AV-A inner and visual-only difference was found in aligned apparent motion condition during 248-298ms. As it was the case for horizontal and vertical connecting bar conditions, the ERP associated with the AV-A outer condition was not found significantly different from the ERP associated with the visual-only condition on aligned apparent motion condition (p>.05) (Figure 3.2B, 3^{rd} row).

AV inner and outer sound on single baseline condition were found significantly different from each other during 42-106 ms, 134-182 ms and 224-266 ms (p<.05). 42-106 ms time interval corresponded to larger P1 component associated with the AV outer sound condition compared to the AV inner sound condition and later 134-182 ms time interval indicated the opposite trend, which is the larger P1 associated with the AV inner sound condition and relatively lower amplitude values associated with the AV outer sound condition. Around 224-266 ms, the difference between earlier N1 component of the AV outer sound condition and larger N1 component of the AV inner sound condition was demonstrated (Figure 3.2A, 4^{th} row). AV-A versus visual-only condition comparison revealed the significant difference between the AV-A inner sound condition and visual-only condition around 264-384 ms (p<.05). The AV-A outer sound condition was not significantly different from the visual-only condition (p>.05) (Figure 3.2B, 4^{th} row).

Lastly, for the single apparent motion with horizontal connecting bar condition, the AV inner and outer sound conditions were found significantly different during 30-106 ms, 138-164 ms and 218-326 ms on frontal channels (p<.05). 30-106 ms time interval corresponded to the larger P1 component of the AV outer sound condition and significantly lower positive voltage change of the AV inner sound condition. 138-164ms time interval, on the other hand, included larger P1 component of the AV inner sound condition and negative voltage change of the AV outer sound condition. The N1 component associated with AV inner sound condition was significantly larger relative to the negative component induced by AV outer sound condition around 218-326 ms (Figure 3.2A, 5^{th} row). Furthermore, the AV-A outer sound condition was found significantly different from the visual-only condition around 306-338 ms and 388-406 ms (p<.05). Besides, the AV-A inner sound condition was found significantly different from the visual-only condition during 24-64 ms and 178-206 ms. Voltage change associated with the visual-only condition was in between the AV-A inner and outer sound conditions although it was significantly different from them at specific time intervals only (Figure 3.2B, 5^{th} row).

Overall, in frontal channels, ERPs associated with AV inner and outer sound condition were found to be significantly different from each other earlier, during the positive components (around 50-100 ms) and later, during the negative components (around 200-300 ms) for all connecting bar conditions. Only 2 aligned apparent motions condition did not indicate the similar trend. Individual effect of inner sound condition was present around early (50- 150 ms) and later time intervals (200-300 ms and 300-400 ms) for all connecting bar conditions except the single baseline condition. For the single baseline condition, individual effect of inner sound condition was observed around early time intervals only and individual effect of outer sound condition was observed after 200 ms only.

Further analysis investigated the AV inner-outer sound differences across connecting bar conditions for the early time interval of 50-100 ms and later time interval of 200-300 ms. Individual paired-samples t-tests indicated that the mean AV inner-outer sound difference was not found significantly different across connecting bar conditions for the early and late time intervals (p<.05). However, a trend indicating larger voltage difference between the sound conditions for the vertical connecting bar condition relative to the horizontal connecting bar condition was observed for both early and later time intervals (Figure 3.3A). Moreover, differences between individual effects of sound conditions were investigated around the 50-150 ms, 200-300 ms and 300-400 ms time intervals. Individual paired-samples t-tests investigating the mean difference between the AV-A inner and AV-A outer sound conditions also did not reveal significant difference across connecting bar conditions for the early and late time intervals (p>.05). Nevertheless, a trend indicating larger voltage difference between the AV-A inner and outer sound conditions was observed for vertical connecting bar condition relative to the horizontal connecting bar condition for the 50-100 ms, 200-300 ms and

300-400 ms time intervals (Figure 3.3B).

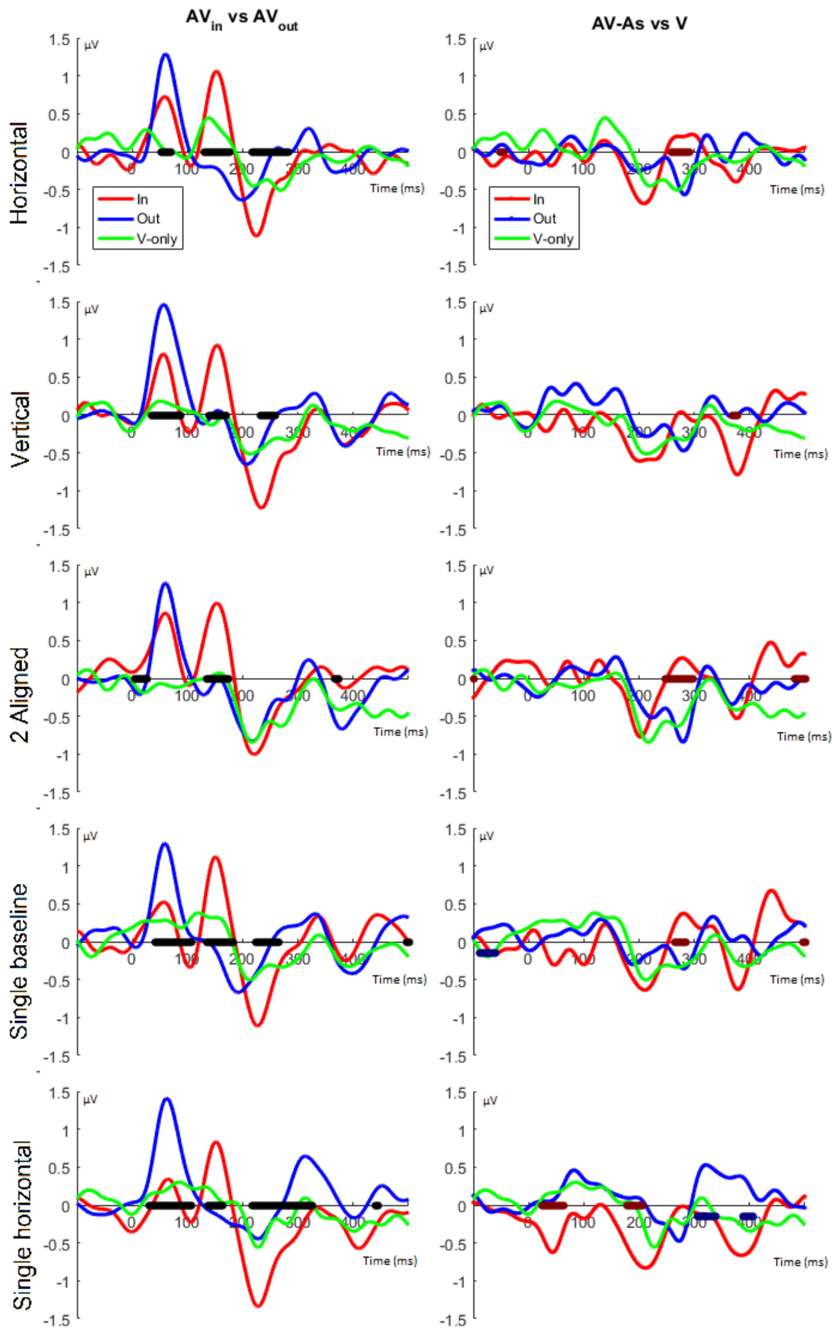

Figure 3.2: Average ERPs from frontal channels (F1, F3, AF3). Inner and outer sound comparisons for five connecting bar conditions. Black lines indicates significant difference interval (A). Audiovisual minus inner-only was compared to visual-only condition (indicated in dark red lines) and audiovisual minus outer-only was compared to visual-only condition (indicated in dark blue lines) (B)

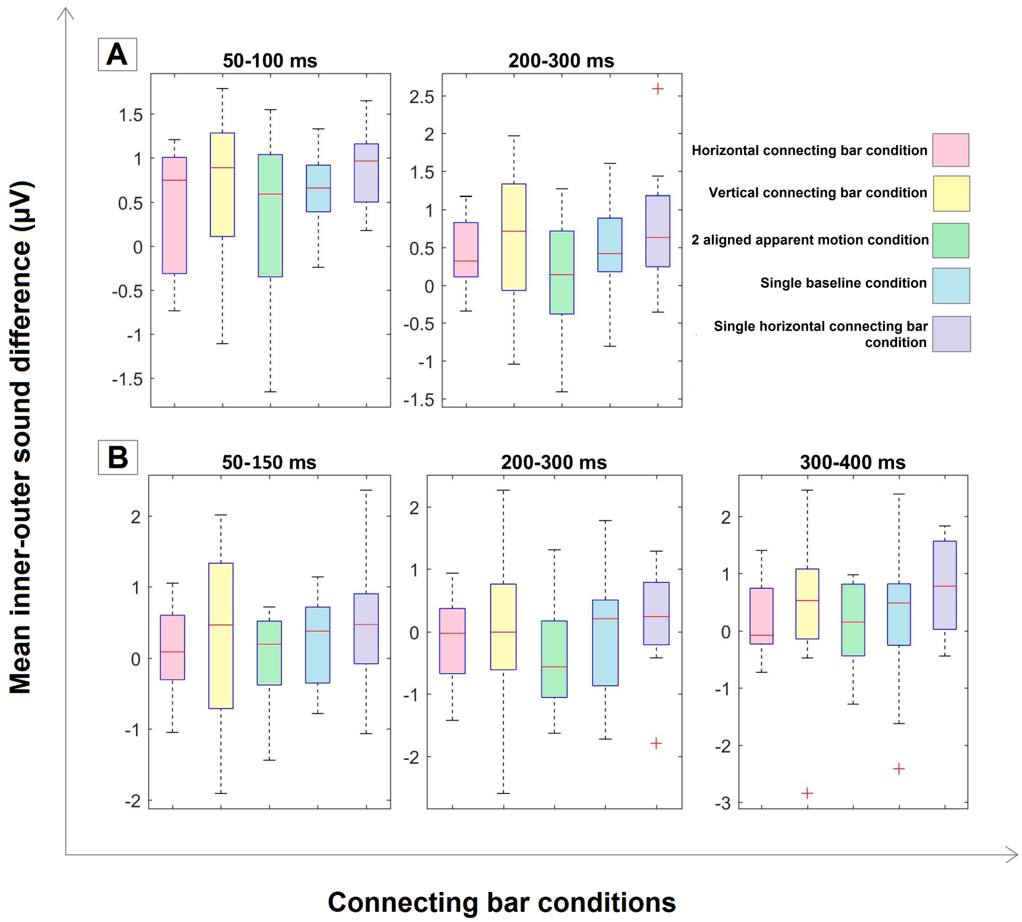

Figure 3.3: Mean AV inner-outer sound differences (A) and mean AV-A inner-outer sound differences (B) across connecting bar conditions for frontal channels. Data points beyond the whiskers are displayed using + sign.

Central Channels

In horizontal connecting bar condition, the AV inner and outer sound conditions were found significantly different from each other around 234-242 ms which corresponded to the slightly positive amplitude values associated with the AV outer sound condition and negative amplitude values associated with the AV inner sound condition after both sound conditions revealed N1s (p<.05) (Figure 3.4A, 1^{st} row). The AV-A inner sound versus visual-only comparison revealed significant differences over 56-78 ms, 246-274 ms and 302-346 ms (p<.05). Time interval of 56-78 ms was around the larger P1 component associated with the AV-A inner sound condition relative to the voltage change associated with the visual-only condition. Larger N1 associated with the visual-only condition was significantly different from the late positive peaks induced by AV-A inner sound condition during 246-274 ms and 302-346 ms (Figure 3.4B, 1^{st} row). Larger P1 component of AV-A outer sound condition was significantly different from the P1 component of visual-only condition around 60-112 ms. The N1 component associated with the visual-only condition was found to be significantly larger in amplitude compared to the N1 component of AV-A outer condition around 142-282 ms ($p<.05$).

For vertical connecting bar condition, statistically significant difference between the AV inner and outer sound conditions were found during 86-130 ms and 242-294 ms ($p<.05$). First interval indicated the difference between positive amplitude values induced by the AV outer sound condition and negative amplitude values induced by the AV inner sound condition. The visual-only condition was observed to maintain its amplitude value in between the two sound conditions. 242-294 ms time interval included the difference between positive amplitude values associated with the AV outer sound condition and negative amplitude values associated with AV inner sound condition when they both returned to near baseline values after revealing N1s. The amplitude of the visual-only condition remained in between the two sound conditions once again (Figure 3.4A, 2^{nd} row). The AV-A outer sound condition was found significantly different from the visual-only condition over 86-108 ms, 176-220 ms and 330-406 ms ($p<.05$). The P1 associated with the AV-A outer sound condition was significantly larger from the P1 component of visual-only condition around 86-108 ms while the N1 associated with the AV-A outer sound condition was significantly smaller compared to the visual-only condition around 176-220 ms. 330-406 ms time interval indicated significantly larger late P1 associated with the AV-A outer sound condition. Moreover, the AV-A inner sound condition was found to be significantly different from the visual-only condition during 18-68 ms ,170-182 ms, 314-352 ms and 414-434 ms ($p<.05$). Similarly, first time interval indicated larger P1, second time interval indicated smaller N1 and the third and fourth time intervals demonstrate late positive peak associated with the AV inner sound condition (Figure 3.4B, 2^{nd} row).

Two aligned apparent motion condition did not indicate a significant difference between the AV inner and outer conditions since both sound conditions revealed similar ERP waves (p>.05) (Figure 3.4A, 3^{rd} row). Nonetheless, the AV-A inner sound condition was found significantly different from the visual-only condition over 32-90 ms, 122-142 ms, 190-226 ms, 242-270 ms and 312-334 ms while the AV-A outer sound condition was found significantly different from the visual-only condition during 72-112 ms, 230-268 ms and 320-394 ms (p<.05). These overlapping time intervals before 150 ms demonstrated the difference between the P1s associated with the sound conditions to the P1 component of visual-only condition. After 200 ms, the N1 associated with visual-only condition was significantly larger from the AV-A sound conditions at mostly overlapping time intervals. N1s of each condition peaked around 150 ms but the sound conditions were not significantly different from the visual-only condition around the peak (p>.05) (Figure 3.4B, 3^{rd} row).

Single baseline condition revealed significant AV inner and outer sound differences around 10-26 ms and after 400 ms (p<.05). ERPs associated with the sound conditions were quite similar. These time intervals may be due to background noise since audiovisual interaction as early as 10 ms and as later as 400 ms were not in accordance with the literature and did not reveal a consistent trend across connecting bar conditions and ROIs (Figure 3.4A, 4^{th} row). The AV-A inner sound condition was found to be significantly different from the visual-only condition during 30-34 ms, 42-86 ms, 106-146 ms, 184-288 ms, 294-352 ms and 404-442 ms .(p<.05). Similarly the AV-A outer sound condition was found to be significantly different from the visual-only condition during 38-150 ms, 184-294 ms and 318-422 ms (p<.05). Individual effects of both sound conditions were found to be significantly different around both early and late time intervals which corresponded to P1, N1 and P2 components of AV-A sound conditions and larger N1 component of visual-only condition (Figure 3.4B, 4^{th} row).

For single horizontal connecting bar condition, the AV inner and outer conditions were not found significantly different from each other (p>.05) (Figure 3.4A, 5^{th} row). However the visual-only condition was significantly different from the

AV-A inner sound condition during 22-174 ms and 202-444 ms and from the AV-A outer sound condition during 38-158 ms, 186-250 ms and 324-354 ms (p<.05). Around 22-174 ms the P1 and N1 of AV-A inner sound and around 38-158 ms the P1 and N1 of AV-A outer sound condition were significantly larger compared to the visual-only condition which maintained negative amplitude values. Around later time intervals, the AV-A sound conditions maintained positive amplitude values and remained significantly different from the negative amplitude values associated with the visual-only condition over 202-444 ms (AV-A inner sound condition) 186-250 ms and 324-354 ms (AV-A outer sound condition). Second negative peak induced by the AV-A outer sound condition, which peaked around 270 ms, was not significantly different from the visual-only condition (p>.05) (Figure 3.4B, 5^{th} row).

In summary, for the central channels AV inner-outer sound differences were observed in both early time intervals (50-100 ms) and later time intervals (200-300 ms) only for vertical connecting bar condition. In horizontal connecting bar condition, the difference in audiovisual condition was observed only around later time intervals (after 200 ms). The AV inner-outer sound conditions in 2 aligned apparent motions, single baseline and single horizontal connecting bar conditions were not found as significantly different from each other. However individual effects of the inner and outer sound conditions were observed around both early and later time intervals for all connecting bar conditions. The effect of the individual sound conditions around the early time intervals was indicated with the difference between the P1s of the AV-A sound and visual-only conditions (50-100 ms) and around later time intervals with the difference between the N1s (200-300 ms) and P2s (300-400 ms) of AV-A sound conditions and visual-only conditions for all connecting bar conditions.

Further analysis investigated the AV inner-outer sound difference between all connecting bar conditions in 50-100 ms and 200-300 ms time intervals. Individual paired samples t-test did not indicate significantly different AV sound difference between the connecting bar conditions for these time intervals (p>.05). However a trend indicating larger AV inner-outer sound difference for the single baseline condition relative to the single horizontal connecting bar condition was observed for both 50-100 ms and 200-300 ms time intervals. The AV sound difference in horizontal and vertical connecting bar conditions did not follow a consistent trend across these time intervals (Figure 3.5A). The difference between the individual effects of each sound condition was compared across all connecting bar conditions around P1s (50-150 ms), N1s (200-300 ms) and P2s (300-400 ms). Individual paired samples t-tests indicated that the AV-A inner-outer difference was not significantly different between the connecting bar conditions around these time intervals ($p > .05$). A trend indicating larger AV-A inner-outer sound difference was observed in single baseline condition and vertical connecting bar condition relative to the single horizontal connecting bar condition and horizontal connecting bar condition respectively (Figure 3.5B).

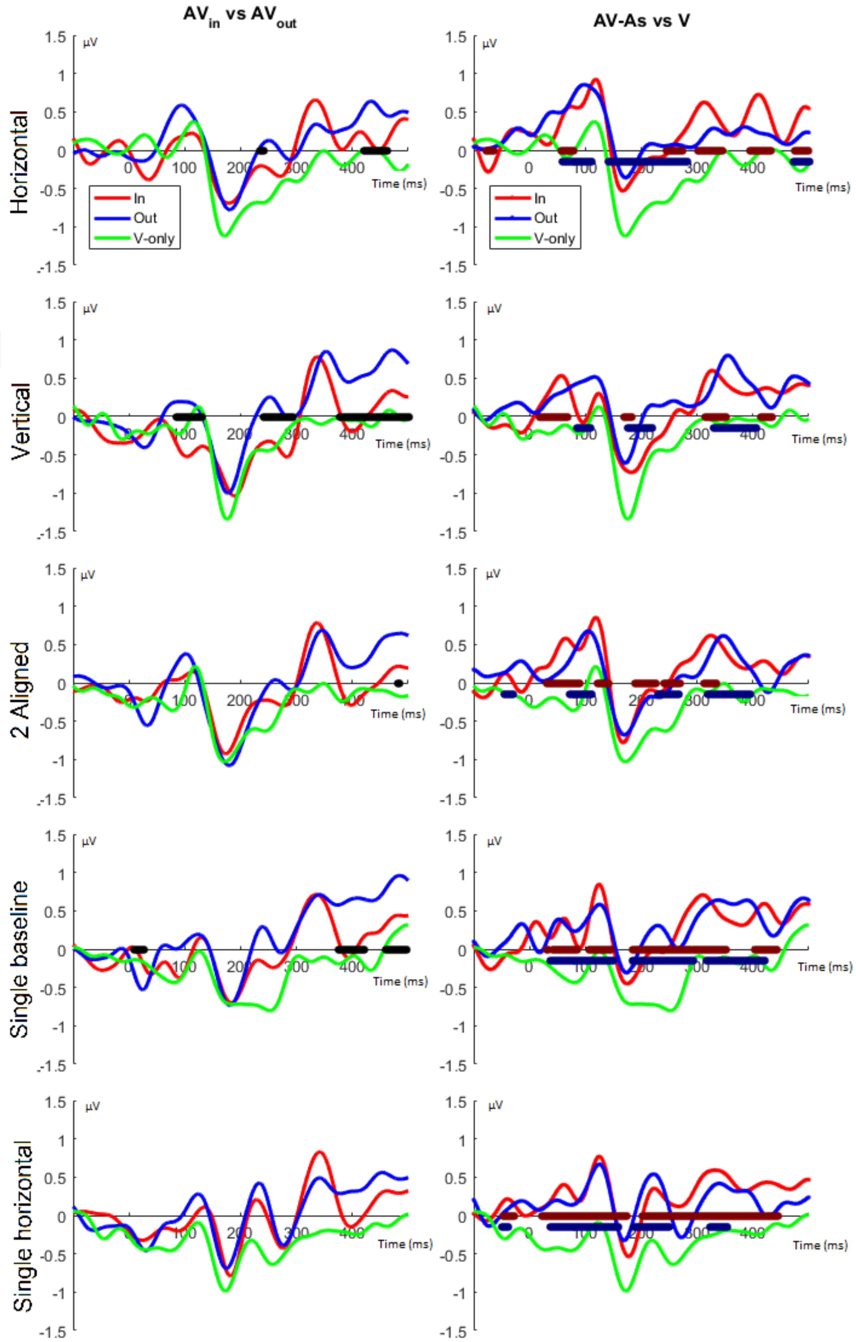

Figure 3.4: Average ERPs from central channels (Cz, C1, C2, C3 and C4). Inner and outer sound comparisons for five connecting bar conditions. Black lines indicates significant difference interval (A). Audiovisual minus inner-only was compared to visual-only condition (indicated in dark red lines) and audiovisual minus outer-only was compared to visual-only condition (indicated in dark blue lines)(B)

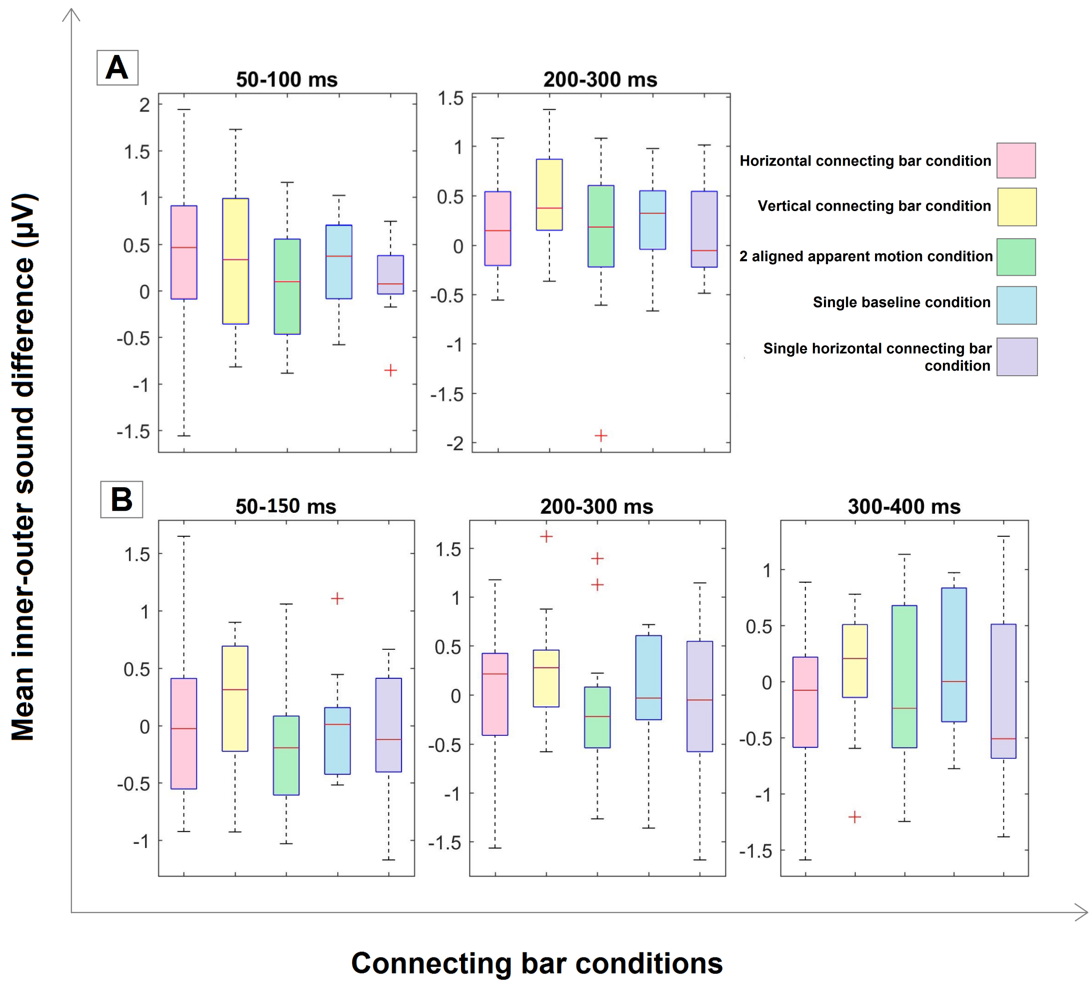

Figure 3.5: Mean AV inner-outer sound differences (A) and mean AV-A inner-outer sound differences (B) across connecting bar conditions for central channels. Data points beyond the whiskers are displayed using + sign.

Temporal Channels

Horizontal connecting bar condition indicated the AV inner-outer sound difference over 32-82 ms, 132-182 ms and 210-274 ms (p<.05). Right after the stimulus onset, the P1 associated with the AV outer sound condition was significantly larger compared to the AV inner sound condition around 32-82 ms. The P1 component of AV inner sound condition was significantly different from the N1 component of the AV outer sound condition during 132-182 ms. Around 210-274 ms time interval, the N1 component of AV inner sound condition was significantly larger relative to the AV outer sound condition which maintained amplitude values around baseline. The visual-only condition revealed an amplitude change between the two sound conditions (Figure 3.6A, 1^{st} row). The AV-A versus visual-only comparison revealed significant difference over 26-44 ms, 254-314 ms and 394-410 ms for the AV-A inner sound condition and 216-250 ms for the AV-A outer sound condition (p<.05). During the stimulus presentation, voltage change associated with the AV-A inner sound condition was significantly lower compared to the visual-only condition. But more importantly after 200 ms, both sound conditions revealed N1s that were significantly different from the N1 component of visual-only condition (Figure 3.6B, 1^{st} row).

On the vertical connecting bar condition, the AV inner-outer difference was found over 42-106 ms, 136-170 ms during the stimulus presentation and over 206-306 ms and 336-404 ms when the stimulus display was disappeared (p<.05). The P1 associated with the AV outer sound condition was found remarkably larger in amplitude compared to the AV inner sound condition around 42-106 ms. The P1 associated with the AV inner sound condition peaked after 100 ms and became significantly larger than the positive amplitude values associated with the AV outer sound condition around 136-170 ms. During 206-336 ms, the N1 component of the AV inner sound condition was significantly different from the N1 associated with the AV outer sound condition which peaked earlier in time and reached positive amplitude values earlier. The visual-only condition followed an amplitude change between the two sound conditions (Figure 3.6A, 2^{nd} row). The AV-A outer sound condition was found to be significantly different from the visual-only condition over 206-246 ms and 352-356 ms (p<.05). The AV-A outer sound condition revealed two small negative components around 200 ms and 300 ms; while the visual-only condition revealed larger negative peak lasted around 150-300 ms. The N1 component of the AV-A outer sound condition was significantly different from the the visual-only condition only when it returned to baseline. Moreover, the P2 component of the AV-A outer sound condition was significantly different from the negative voltage change associated with the visual-only condition. The AV-A inner sound condition indicated similar pattern as the visual-only condition. Thus, it was not found significantly different from the ERP associated with the visual-only condition (p>.05) (Figure 3.6B, 2^{nd} row).

2 aligned apparent motions condition indicated significant AV inner-outer sound difference during 48-86 ms, 136-178 ms and 232-248 ms (p<.05). The P1 associated with the AV outer sound condition was observed to be larger in amplitude compared to positive amplitude values of AV inner sound condition around 48-86 ms while the P1 associated with the AV inner sound condition was found to be larger compared to the positive voltage values associated with the AV outer sound condition around 136-178 ms. Also the N1 associated with the AV inner condition was observed to have larger amplitude compared to the amplitude changes associated with the AV outer sound condition around 232-248 ms (Figure 3.6A, 3^{rd} row). N1 of the AV-A inner sound condition peaked around 200 ms and found to be significantly different from the visual-only condition around 240-286 ms. N1 associated with the AV-A outer sound condition was significantly different from the larger N1 component of the visual-only condition around 228-254 ms (p<.05) (Figure 3.6B, 3^{rd} row).

For the single baseline condition, the AV inner sound condition was found to be significantly different from the AV outer sound condition over 46-68 ms, 89-96 ms and 130-184 ms during stimulus display and 228-264 ms, 388-396 ms after stimulus display (p<.05). The P1 associated with the AV outer sound condition was significantly larger compared to the positive voltage values associated with the AV inner sound condition over 46-68 ms and 89-96 ms. Around 130-184 ms, P1 associated with the AV inner sound condition was found to be larger compared to the positive amplitude values associated with the AV outer sound condition. During 228-264 ms, the AV outer condition revealed earlier N1 component and arrived at positive amplitude values earlier than the AV inner condition. The voltage difference around 388-396 ms corresponded to the end of the P2 components associated with both sound conditions (Figure 3.6A, 4^{th} row). Smaller N1 component of the AV-A outer condition was found significantly different from the N1 component of the visual-only condition during 192-248 ms and the P2 associated with the AV-A outer sound condition was significantly different from the negative amplitude values associated with the visual-only condition around 320-352 ms and 388-390 ms. The N1 associated with the AV-A inner condition was significantly different from the N1 associated with the visual-only condition over 252-278 ms (p<.05) (Figure 3.6B, 4^{th} row).

On the single horizontal connecting bar condition, the difference between the AV inner and outer sound conditions were found significant over 44-108 ms, 140-168 ms, 208-268 ms, 298-352 ms and 384-414 ms (p<.05). The P1 associated with the AV outer sound condition was found significantly larger from the voltage change associated with the AV inner sound condition around 44-108 ms. Around 140-168 ms, P1 associated with the AV inner condition was found significantly larger from the voltage change associated with the AV outer sound condition. Between 208 to 268 ms, the AV outer sound condition arrived at positive amplitude values after the N1 while the AV inner condition maintained its significantly larger N1 component. Moreover, P2s of the AV inner and outer sound conditions were found significantly different from each other during 298-352 ms (Figure 3.6A, 5^{th} row). The AV-A outer sound condition was significantly different from the visual-only condition during 222-246 ms, 320-348 ms and 384-402 ms (p<.05). After 200 ms, the voltage change associated with the visual-only condition was in between two sound conditions but it was significantly different from only the AV-A outer sound condition. Moreover the P2 associated with the AV-A outer sound condition was significantly larger than the ERP of the the visual-only condition around 320-348 ms and 384-402 ms (Figure 3.6B, 5^{th} row). The AV-A inner sound condition was significantly different from the visual-only condition during 268-278 ms, which indicated the larger N1 component of the visual-only condition relative to the AV-A inner sound condition (p<.05).

Overall, in temporal channels AV sound differences were observed both early and later time intervals for all connecting bar conditions. The early time intervals corresponded to the difference between the P1s associated with the AV sound conditions (50-100 ms) while the late time intervals corresponded to the difference between N1s and late positive components associated with the AV sound conditions (200-300 ms). Individual effects of inner and outer sound conditions were observed around later time intervals for all connecting bar conditions.

The difference between the AV inner-outer sound conditions were investigated between the connecting bar conditions for 50-100 ms and 200-300 ms time intervals. Individual paired-samples t-tests indicated that the AV sound difference was not significantly different between the connecting bar conditions for both the early and later time intervals ($p>.05$). Nevertheless, a trend indicating larger AV sound difference for the vertical connecting bar condition relative to the horizontal connecting bar condition was observed for both time intervals. For the single apparent motion condition, larger AV sound difference was observed for single horizontal connecting bar condition relative to the single baseline condition (Figure 3.7A). The difference between individual effects of inner and outer sound conditions was compared between the connecting bar conditions for 50-150 ms, 200-300 ms and 300-400 ms time intervals. The AV-A sound difference was not significantly different between the connecting bar conditions, but a trend indicating larger AV-A inner-outer sound difference for the vertical connecting bar condition relative to the horizontal connecting bar condition was observed. A consistent trend for the single connecting bar conditions was not observed (Figure 3.7B).

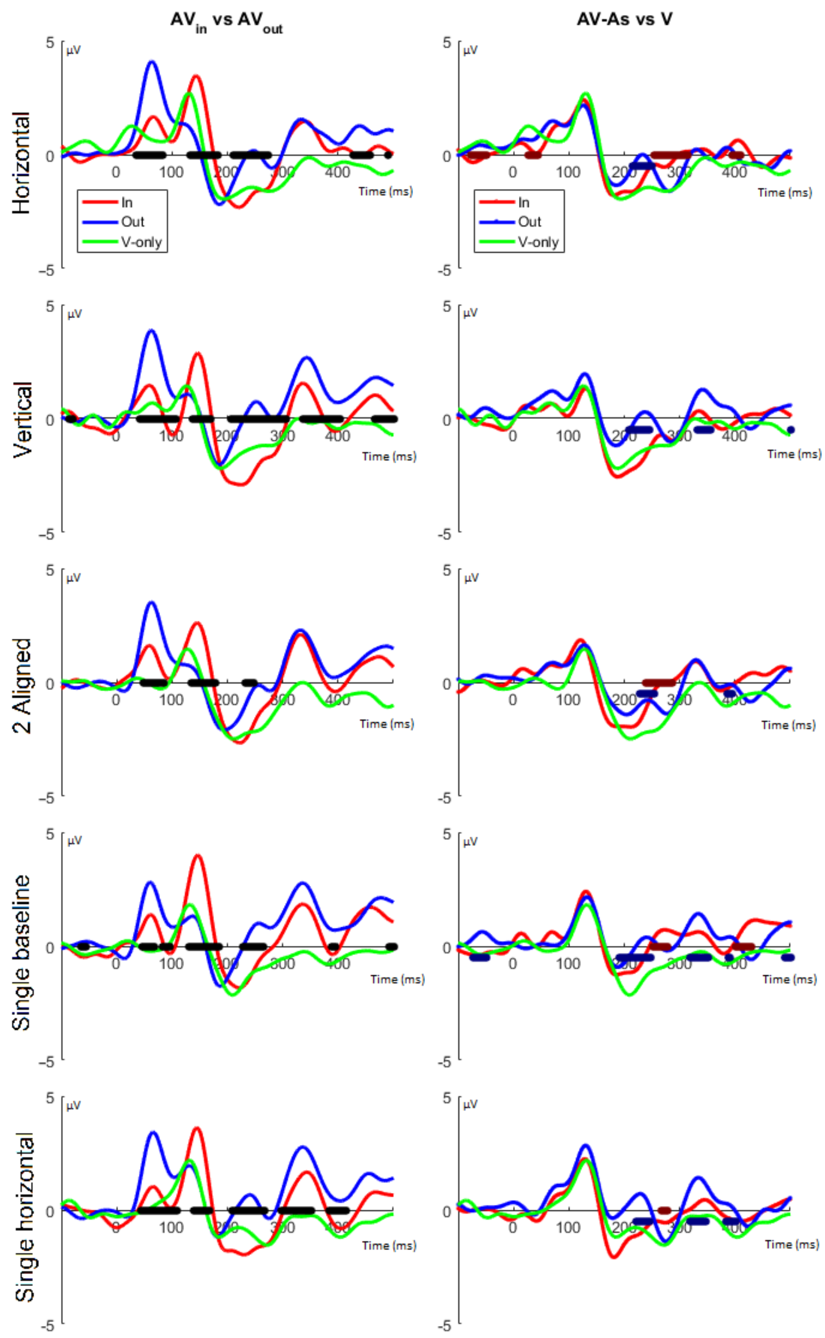

Figure 3.6: Average ERPs from temporal channels (T7, TP9, T8 and TP10). Inner and outer sound comparisons for five connecting bar conditions. Black lines indicates significant difference interval (A). Audiovisual minus inner-only was compared to visual-only condition (indicated in dark red lines) and audiovisual minus outer-only was compared to visual-only condition (indicated in dark blue lines)(B).

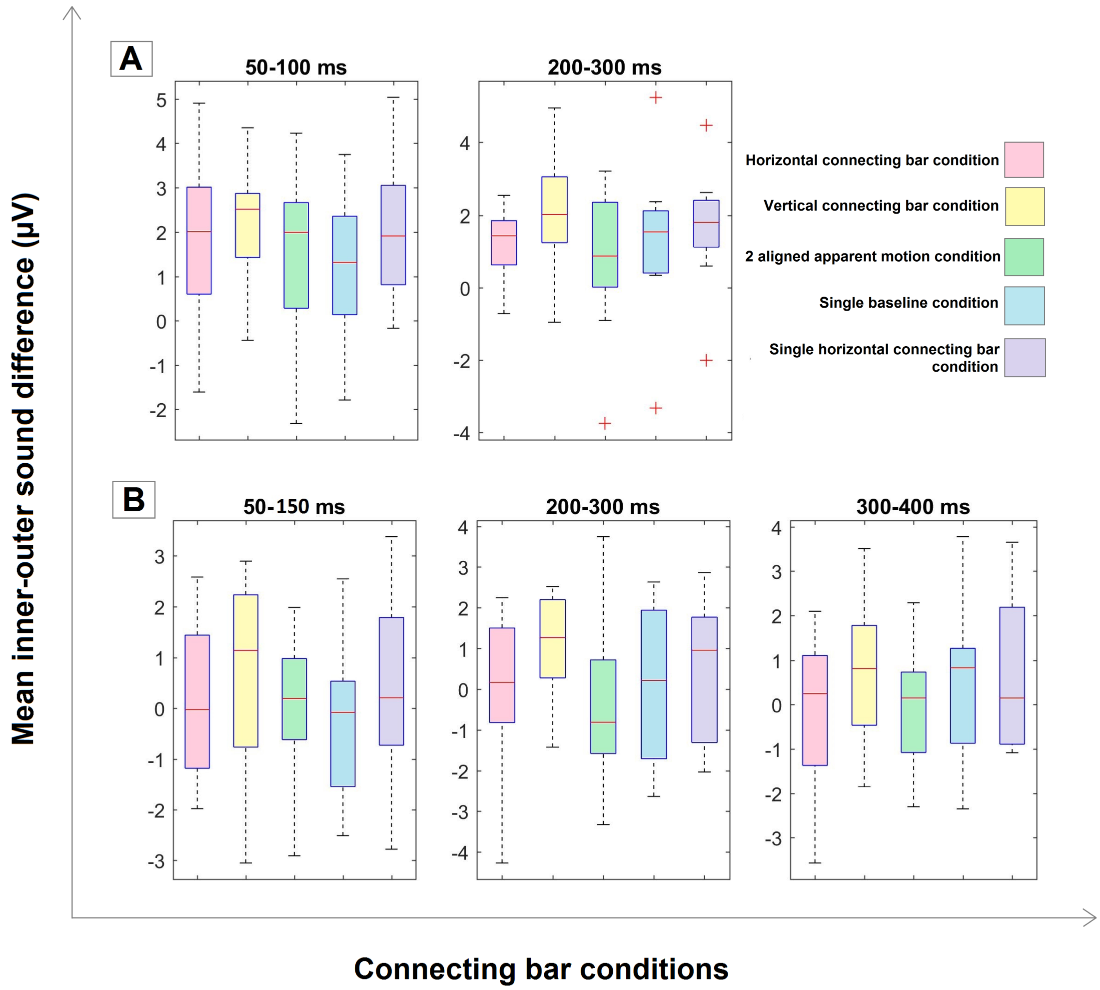

Figure 3.7: Mean AV inner-outer sound differences (A) and mean AV-A inner-outer sound differences (B) across connecting bar conditions for temporal channels. Data points beyond the whiskers are displayed using + sign.

Occipital-Parietal Channels

Horizontal connecting bar condition revealed statistically significant AV inner and outer sound difference during 32-98 ms, 142-164 ms, 216-278 ms and 422-464 ms (p<.05). The P1s of the AV inner and outer sound conditions were significantly different from each other around 32-98 ms, which was during the stimulus presentation. Significantly larger N1 associated with the AV outer sound condition was observed around 142-164 ms. After the stimulus presentation, the AV inner and outer sound conditions was significantly different from each other during 216-278 ms. Around this interval, the AV outer sound condition revealed positive amplitude values while the AV inner sound condition maintained its negative amplitude values after the peak of N1 (Figure 3.8A, 1^{st} row). During stimulus presentation, the AV-A inner sound condition was significantly different from the visual-only condition during 56-76 ms while the AV-A outer wave was significantly different from the visual-only condition for a larger time interval: 56-106 ms and 118-186 ms (p<.05). These distinct intervals found for the AV-A inner and outer conditions around P1s, although the ERPs of both sound condition look similar, might be the consequence of higher inter-subject variation in the AV-A inner condition. After 200 ms, both sound conditions arrived at positive amplitude values and became significantly different from the visual-only condition, specifically during 192-322 ms (AV-A inner) and 210-274 ms (AV-A outer) (p<.05) (Figure 3.8B, 1^{st} row).

On vertical connecting bar condition, the AV inner and outer conditions were found significantly different from each other during 88-116 ms, 150-170 ms and 202-292 ms (p<.05)(Figure 3.8A, 2^{nd} row). The P1 associated with the AV outer sound condition and slightly negative amplitude values associated with the AV inner sound condition was significantly different from each other during 88-116 ms. The N1s associated with both AV sound conditions were significantly different from each other around 150-170 ms. Positive amplitude values associated with the AV outer sound condition was significantly different from the negative amplitude values associated with AV inner sound condition around 202-292 ms. Similar trend was observed in horizontal connecting bar condition too. However the significant difference interval between the AV inner and outer sound conditions was found to be larger than it was for horizontal connecting bar condition. The AV-A inner sound condition was found to be significantly different from the visual-only condition during 26-70 ms, 176-196 ms, 252-272 ms and 318-362 ms (p<.05). On the other hand, the AV-A outer condition was found to be significantly different from the visual-only condition over 30-52 ms, 106-132 ms, 176-230 ms and 334-394 ms (p<.05). The AV-A inner and outer differences from the visual-only condition mostly overlap on occipital channels. Before the P1s peaked around 100 ms, both sound conditions were found to have significantly larger amplitude values compared to the visual-only condition, which maintained near baseline values. However only the P1 associated with the AV-A outer sound condition was found to be significantly larger compared to the P1 component of the visual-only condition. Around 176- 196 ms, the N1 component of the AV-A inner sound condition and around 176- 230 ms the N1 component of the AV-A outer sound conditions were found to be significantly smaller compared to the N1 component of the visual-only condition. After 200 ms, both sound conditions indicated significantly larger P2s compared to the visual-only condition (Figure 3.8B, 2^{nd} row).

For the 2 aligned apparent motions condition, the difference between the AV inner and outer conditions were found to be significant around 68-98 ms and 126-156 ms which was during the stimulus presentation ($p<.05$)(Figure 3.8A, 3^{rd} row). The difference between the P1 components of both sound conditions were significant around 68-98 ms and the N1 component of the AV inner sound condition was significantly smaller relative to the N1 component of AV outer sound condition around 126-156 ms. ERP of the AV-A inner condition was found to be significantly different from the ERP of the visual-only condition during 30-32 ms, 64-86 ms, 198-218 ms, 250-276 ms and 308-330 ms while the ERP of the AV-A outer sound condition was found significantly different from the ERP of the visual-only condition during 86-112 ms and 320-372 ms ($p<.05$). The P1 component of the visual-only condition was significantly smaller relative to the P1 component of AV-A inner sound condition around 30-32 ms, 64-86 ms time interval and relative to the P1 component of AV-A outer sound condition around 86-112 ms time interval. N1 component of the visual-only condition was only significantly different from the N1 component of the AV-A inner sound condition (around 198-218 ms). During 250-276 ms and 308-330 ms, the P2 component of the AV-A inner sound condition was larger compared to the negative amplitude values associated with the visual-only condition. Over 320-372 ms, significantly larger P2 component of the AV-A outer sound condition compared to the visual-only condition was observed (Figure 3.8B, 3^{rd} row).

Single baseline condition indicated the AV inner and outer difference around 16-32 ms, 82-92 ms, 144-156 ms and 214-274 ms ($p<.05$)(Figure 3.8A, 4^{th} row).

Around 16-32 ms the AV outer sound condition induced a negative component while the AV inner sound condition was near baseline. Around 82-92 ms it was the other way around. 144-156 ms time interval corresponded to the difference between the AV sound conditions around the end of the P1s and in the beginning of the N1s. The AV-A versus visual-only comparison revealed highly overlapped significant difference intervals for the inner and outer sound conditions. The AV-A inner sound condition was found to be significantly different from the visual-only condition over 34-140 ms, 188-370 ms and 402-442 ms while the AV-A outer condition was found to be significantly different from the visual-only condition during 54-154 ms, 180-284 ms, 328-396 ms and 412-428 ms (p<.05). These overlapping significant difference intervals indicated that both sound conditions had an equal effect on single baseline condition (Figure 3.8B, 4^{th} row). P1 of the AV-A inner sound condition was significantly larger from the P1 of the visual-only condition around 34-140 ms and P1 of AV-A outer sound condition was significantly larger from the P1 of visual-only condition around 54-154 ms time interval. The N1 and P2 associated with the AV-A inner sound condition was significantly different from the larger N1 associated with the visual-only condition around 188-370 ms and 402-442 ms. Similarly the N1 and P2 associated with the AV-A outer sound condition was significantly different from the N1 and P2 components of the visual-only condition around 180-284 ms, 328-396 ms and 412-428 ms time intervals.

Finally on the single horizontal connecting bar condition, significant AV inner and outer difference was observed during 216-268 ms only (p<.05). Late positive peak associated with the AV outer sound condition was found to be significantly larger than second negative component of the AV inner condition (Figure 3.8A, 5^{th} row). The AV-A inner sound condition was significantly different from the visual-only condition during 36-88 ms and 254-362 ms while the AV-A outer sound condition was significantly different from the visual-only condition during 48-64 ms, 106-152 ms, 206-238 ms and 326-352 ms (p<.05). In the beginning of the stimulus presentation, both ERPs associated with the AV-A inner (36-88 ms) and AV-A outer (48-64 ms) conditions was found to be significantly different from the visual only condition. Both sound conditions induced positive amplitude values while the visual-only condition indicated negative amplitude values. Around 106-152 ms, P1 component of the AV-A outer condition was found to be significantly larger than the P1 component of the visual-only condition. When the stimulus disappeared, P2 component of the AV-A outer condition was found significantly larger from the visual-only condition around 206-238 ms (p<.05). P2 component of the AV-A inner sound condition was significantly different from the N1 component of the visual-only condition later in time, around 254-362ms (p<.05) (Figure 3.8B, , 5^{th} row).

To sum up, in occipital-parietal channels individual effects of inner and outer sound conditions were observed around both early (50-150 ms) and later (200-300 ms and 300-400 ms) time intervals. However AV sound differences were observed for only horizontal and vertical connecting bar conditions around both early (50-100 ms) and later (200-300 ms) time intervals. 2 aligned apparent motion condition only indicated AV sound difference at earlier time intervals only while single horizontal connecting bar condition indicated an AV sound difference at later time intervals. The reasons for the audiovisual interaction earlier in time for multiple apparent motions without the connecting bars and later in time for the single apparent motion with horizontal connecting bars are unknown. Individual t-tests between connecting bar conditions on both early and late time intervals may explain the observed trend.

The mean difference between the AV inner-outer sound condition was investigated across all connecting bar conditions around 50-100 ms and 200-300 ms time intervals. Individual paired samples t-tests indicated that AV sound difference was not significantly different between any of the connecting bar conditions around both time intervals (p>.05). Larger AV sound difference for single baseline condition and vertical connecting bar condition was observed only around 200-300 ms, thus a consistent trend was not observed (Figure 3.9A). The mean difference between the individual effects of each sound condition was investigated around 50-150 ms, 200-300 ms and 300-400 ms time intervals. Individual paired-samples t-test did not indicate significantly different AV-A inner-outer sound difference across connecting bar conditions (p>.05). A trend indicating larger

AV-A sound difference in the vertical connecting bar condition relative to the horizontal connecting bar condition was observed. Similar trend for the single baseline condition was observed only around late time intervals (Figure 3.9B).

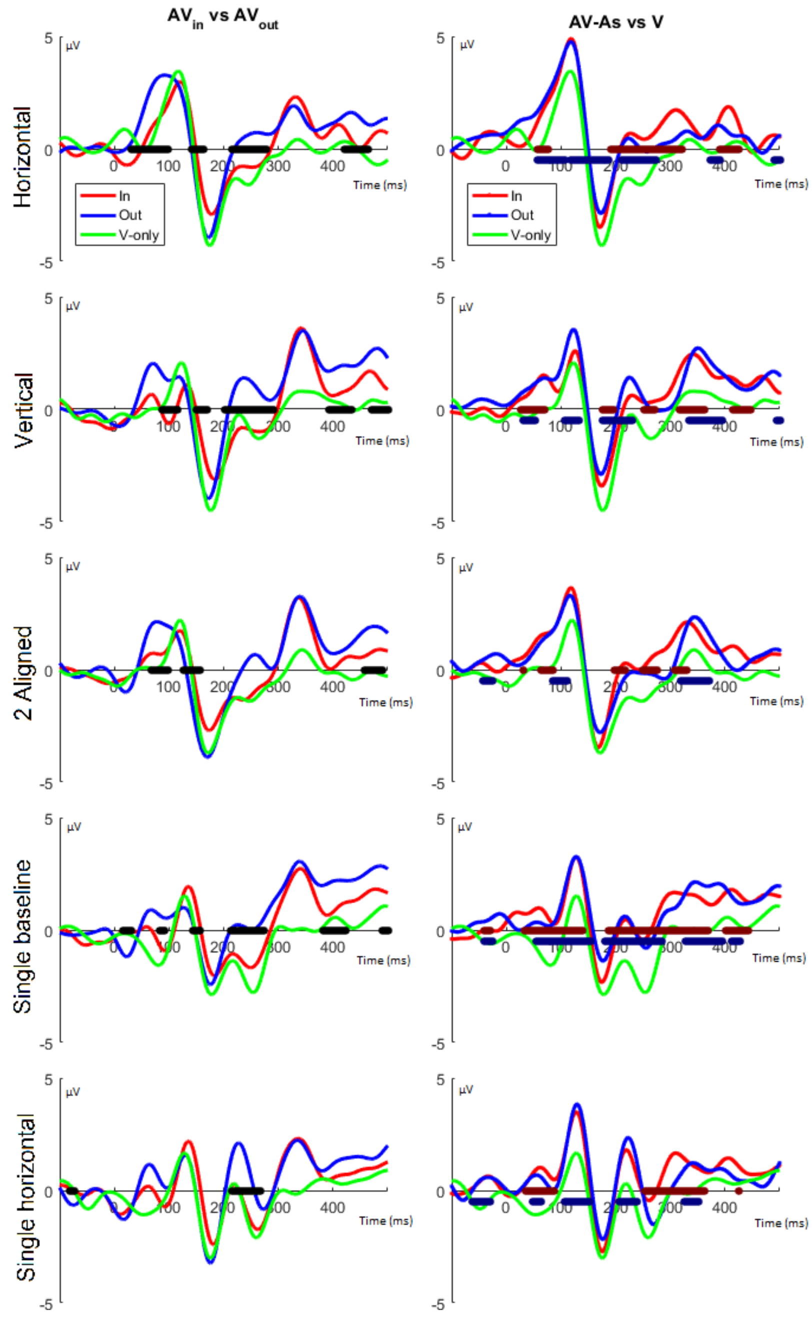

Figure 3.8: Average ERPs from occipital-parietal channels (Pz, PO3 and PO4). Inner and outer sound comparisons for five visual grouping conditions. Black lines indicates significant difference interval (A). Audiovisual minus inner-only was compared to the visual-only condition (indicated in dark red lines) and audiovisual minus outer-only was compared to the visual-only condition (indicated in dark blue lines) (B)

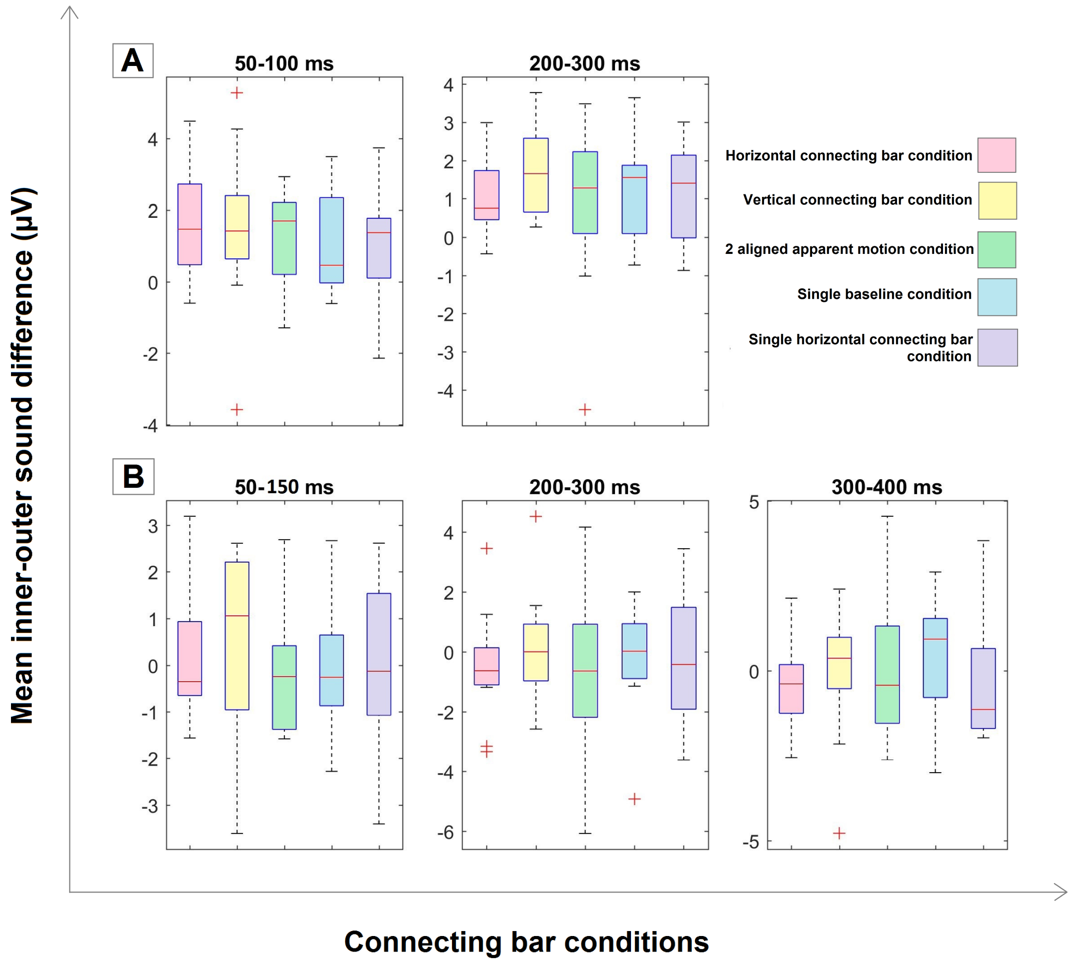

Figure 3.9: Mean AV inner-outer sound differences (A) and mean AV-A inner-outer sound differences (B) across connecting bar conditions for occipital-parietal channels. Data points beyond the whiskers are displayed using + sign.

Overall, AV and AV-A inner-outer sound differences were observed around both early and later time intervals. Individual paired-samples t-test between the connecting bar conditions for each ROI did not reveal significantly different AV and AV-A sound differences between the connecting bar conditions. A consistent trend indicating larger AV inner-outer difference for the vertical connecting bar condition relative to the horizontal connecting bar condition was observed in the temporal and frontal channels while larger AV inner-outer difference for the single baseline condition relative to the single horizontal connecting bar condition was observed in the central channels. Moreover, a trend indicating larger difference between the individual effects of the sound conditions for the vertical connecting bar condition relative to the horizontal connecting bar condition was observed across all ROIs while the similar trend was observed for the single baseline condition relative to the single horizontal connecting bar condition only in the central channels.

3.3 Behavioral Data Obtained from the EEG Session

Behavioral data was analyzed with individual paired-samples t-tests between each connecting bar condition. Only the single horizontal connecting bar and single baseline conditions were found significantly different from each other ($t(11)$=-3.108, p=01) (Bonferroni corrected). Single baseline condition was found to be influenced from the accompanying sounds more than the single apparent motion with horizontal connecting bar condition (Figure 3.10A). This result was in accordance with the findings of Experiment 3. The horizontal connecting bar introduced between the flashes grouped them and created a final percept of a single object moving. Therefore the clicks of each sound condition could not effectively bind to these flashes, which were not perceived independently but as parts of a single object, and the effect of auditory capture decreased.

The two apparent motions grouped with horizontal and vertical connecting bars did not significantly differ from each other (p>.05). However the observed trend was similar to the one indicated in the Experiment 2. Inner sound condition was perceived as faster for the horizontal connecting bar condition more than for the vertical connecting bar and 2 aligned apparent motions conditions. The vertical connecting bar and 2 aligned apparent motions conditions were equally influenced by the accompanying sounds (Figure 3.10B). While the horizontal connecting bar created a final percept of two separate objects moving in left-right direction, vertical connecting bar created a final percept of a single bigger apparent motion. Preceding and following flashes of the single bigger apparent motion was succesfully captured by the clicks of auditory conditions; however the effect of clicks over the multiple first and second flashes was relatively less.

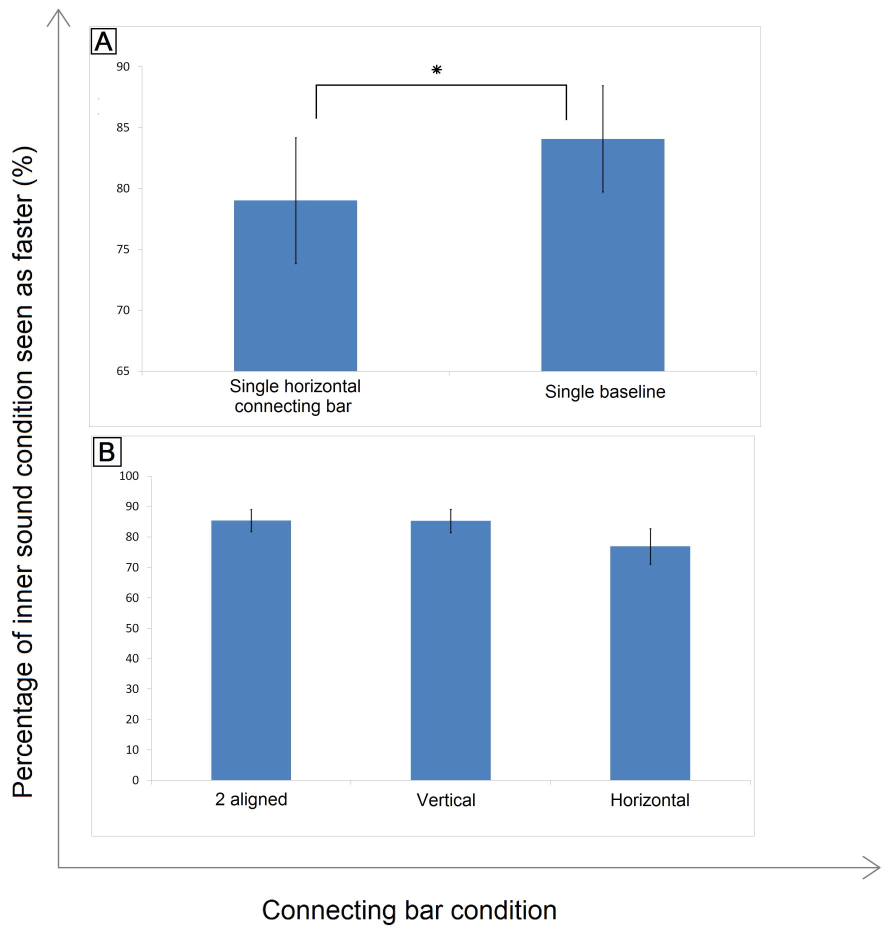

Figure 3.10: Percentage of inner sound seen as faster across single apparent motion with horizontal connecting bar and single baseline conditions (A), 2 aligned apparent motions, vertical connecting bar and horizontal connecting bar conditions (B). Error bars indicate ± SEM.

Chapter 4

General Discussion

4.1 Summary of Findings

4.1.1 Behavioral Results

Auditory stimuli have been shown to dominate over visual stimuli in case of judgments regarding temporal information such as rate of stimulus presentation, time of onset and speed. Previous work suggested that brief sounds introduced before and after the visual stimuli may change the timing of visual stimuli and increase the temporal segregation between them [4, 16, 7]. In the current study, a similar capturing effect of sound on visual stimuli in the temporal domain was studied by using apparent motion stimuli under different gestalt spatial organization principles. Experiment 1 employed two vertically placed apparent motions grouped with UC principle (horizontal and vertical connecting bars), proximity principle (vertical spatial distances between each apparent motion) and common fate principle (each apparent motion was moving to different directions). The results indicated that auditory stimuli modulated participants' speed judgment the most when the apparent motions were grouped with UC principle. The vertical distance between the apparent motions did not directly affect audiovisual interaction. However, the significant interaction effect between the connecting bar condition and spatial proximity conditions suggested that spatial organizations formed with different connecting bar conditions had an effect on the auditory capture depending on the levels of proximity between apparent motions. More specifically, the horizontal and vertical connecting bars had a significant effect between the 2.7°and 4.4°levels of spatial proximity condition, but not on 1.5°. Although none of the connecting bar conditions followed a consistent trend across vertical distance parameters, overall less percentage of inner sound condition seen as faster observed on horizontal connecting bar condition relative to the vertical connecting bar and the 2 aligned apparent motions conditions was in accordance with our expectations of less auditory capture when the number of stimuli from visual and auditory modalities do not match. In other words, horizontal connecting bar created a perception of 2 distinct objects and the inner sound or outer sound conditions could not efficiently captured the both. Furthermore, the common fate condition indicated a trend of decreasing percentage of inner sound condition seen as faster with increasing spatial proximity between each apparent motion, even though this decrease did not lead to a significant effect. The common fate and 2 aligned apparent motion conditions did not follow a similar pattern across the spatial proximity conditions. Nevertheless, auditory capture over the apparent motion moving to different directions was found to be significantly larger in the smallest spatial proximity condition relative to the largest. Sanabria and colleagues [41] previously reported that gestalt grouping principles applied to visual stimuli were found to be facilitating cross-modal binding. Experiment 1 demonstrated that out of three different gestalt grouping principles, only apparent motions grouped with uniform connectedness principle significantly influenced audiovisual interaction, not the apparent motion grouped with common fate and spatial proximity principles. This result supports the findings of Palmer and Rock [9] which suggested that uniform connectedness grouping principle is more fundamental in the visual system. Since multiple objects are connected with physical solid lines, they are represented at the retina as one and they enter the visual system as a single object. Therefore, experiment 2 focused only on the uniform connectedness principle while keeping spatial proximity and common fate parameters constant. Multiple apparent motions were placed vertically with 2.7°spatial proximity and each apparent motion moved in the same direction (left-right).

Experiment 2 was designed in order to investigate the effect of uniform connectedness principle in the audiovisual interaction in case of more than 2 visual stimuli. Furthermore, experiment 2 used variable ISI value between the flashes of apparent motion stimuli in order to observe the effect of sound conditions on physically faster and slower apparent motions. It was demonstrated that inner sound condition shifted the ISI between flashes towards smaller values and generated a perception of faster apparent motion and outer sound condition shifted the ISI between flashes towards larger values and generated a perception of slower apparent motion than it physically was. This modulatory effect of inner and outer sound conditions was larger in the vertical connecting bar condition relative to the horizontal connecting bar condition. Because, similar to the first experiment, vertical connecting bars generated a perception of single apparent motion moving left to right or right to left while horizontal connecting bars generated a perception of 3 distinct apparent motions. Sounds could not efficiently captured the individual visual elements in horizontal connecting bar condition. One interesting finding from Experiment 2 was significantly different effect of the inner sound condition between horizontal and vertical connecting bar conditions. Although the effect of the outer sound condition on the vertical connecting bar condition was found greater compared to the inner sound condition, it was not significantly different from the effect of the outer sound condition on horizontal connecting bar condition. The reason for these distinct effect of the inner sound condition on connectin bar conditions might be different perceptual latencies of visual and auditory stimuli. It was suggested that the neural processing of visual information takes more time relative to the auditory processing [44].Ohmura [45] took account of different perceptual latencies of the auditory and visual systems by presenting the sound in the midway of the flashes and reported enhanced audiovisual interaction. Thus, in the current experiment, the inner sound condition may have modulated the visual conditions more than the outer sound condition.

Experiment 3 was designed to investigate the effect of sound on the speed judgment of single apparent motion grouped with connecting bars. I was demonstrated that the horizontal connecting bar between the flashes of single apparent motion created a clear illusion of apparent motion moving in left-right direction, as it was previously reported by Shepard and colleagues [10]. In accordance with the assumption of unity, two static clicks did not efficiently bind to the single apparent motion since the connecting bar between the flashes created a perception of single continuous object and the auditory capture over single apparent motion with connecting bar critically decreased.

Taken together, the results from the three experiments were in accordance with the effect of static sounds on apparent motion reported by Freeman and Driver [16] and Kafaligonul and Stoner [7]. The current study extended these findings by suggesting an effective intra-modal grouping, uniform connectedness, principle in order to facilitate inter-modal pairing.

4.1.2 EEG Results

Crossmodal interactions have been thought to occur later in information processing hierarchy such as frontal, temporal and parietal regions [46]. Nonetheless, both animal studies and EEG studies indicated auditory stimuli processed in occipital cortex [47, 40]. In our EEG experiment, individual t-tests across 63 channels revealed significance in both the higher order and the occipital regions. Averaged ERPs from the frontal, central, temporal and occipital-parietal channels were investigated across all connecting bar conditions.

In the current study, the difference between P1s associated with the audiovisual conditions was observed around 50-100 ms time interval and the difference between the N1s associated with audiovisual conditions was observed around 200-300 ms time interval. These time intervals were marked as the audiovisual interaction points. AV-A analysis indicated the diffences in P1s associated with the individual effects of the inner and outer sound conditions around 50-150 ms. The differences of N1s associated with the individual effects of the inner and outer sound conditons were around 200-300 ms. Additionally, the P2s associated with the individual effects of sound conditions were found to be significantly different from visual-only conditions around 300-400 ms time interval and this late positive activity was marked as an audiovisual interaction point too. Similar late positive peaks were previously reported in parietal and occipital regions [48, 49]. However, Yang and colleagues observed an audiovisual interaction pattern around 300-340 ms on frontal and central regions and further reported that this audiovisual interaction patterns have disappeared with the increasing sound frequency [49]. Central-frontal positivity was considered to be related with auditory attention and found to be enhanced when the subjects attended to the auditory stimulation [50].

N1 component around 160-180 ms in response to the audiovisual stimulus was previously reported over frontal, central and occipital channels [34, 35, 51]. Talsma and colleagues suggested that N1 in response to audiovisual stimuli on central, parietal an occipital channels was equal to N1 associated with visual-only condition when its auditory component was subtracted [51]. Therefore audiovisual processing was considered the sum of the sensory processing of each modality involved. However, current results indicated that N1 associated with the AV inner and AV outer conditions were significantly different from the visual-only condition after their auditory components were subtracted, especially on central, temporal and occipital-parietal regions. Moreover, early P1s around 100 ms on both anterior and posterior channels in response to the audiovisual stimuli, which were also reported previously [32, 34, 52, 53], were significantly different from the P1s associated with the visual-only conditions after their auditory components were subtracted. These results demonstrated that audiovisual processing is not a simple summation of activity associated with the unimodal visual and auditory components as the additive model suggested [39]. In fact, the P1 and P2 components of AV-A sound conditions were found to be larger than the P1s and P2s of the visual-only condition in the cental and occipital-parietal channels. This trend could fit to the supra-additive model, which suggest that audiovisual condition reveals a component with a larger amplitude relative to the sum of its unisensory components. More specifically, our results indicating larger amplitude values of AV-A condition relative to the visual only condition is consistent with the larger amplitude value of AV condition relative to the sum of its auditory and visual components. On the other hand, N1 components of the AV-A sound conditions were found to be smaller in amplitude relative to the N1 components of the visual-only condition. This trend fits to te sub-additive model which suggest that audiovisual component is smaller in amplitude relative to the sum of its unisensory components [34]. Overall, these interaction sites could not fit to one model but around each time interval the AV-A component was significantly different than the visual-only condition. It is possible to conclude that audiovisual interaction is an individual non-linear process, rather than the sum of its unisensory components.

The ERPs associated with audiovisual inner and outer sound conditions were significantly different for each connecting bar conditions. However, the mean difference between the AV and AV-A inner and outer sound conditions was not significantly different across connecting bar conditions. Nevertheless, a trend indicating larger mean AV inner-outer sound difference for the vertical connecting bar condition relative to the horizontal connecting bar condition was observed over the frontal and temporal channels in both early and late time intervals and over the central and occipital-parietal channels around only late time intervals. The mean difference between the AV-A inner and outer sound conditions were larger for the vertical connecting bar condition relative to the horizontal connecting bar condition across all ROIs which suggested that individual effects of each sound condition tend to be larger when the flashes were grouped with vertical connecting bars and perceived as a single object.

Moreover, the mean AV inner-outer difference and the difference between individual effects of the sound conditions were consistently greater in the single baseline condition for only the central channels. The occipital-parietal channels revealed a similar trend around the later time intervals (N1s and P2s). Depending on the behavioral experiments, it was expected that ERP associated with single apparent motion without a connecting bar between its flashes would indicate greater sound difference, since these two flashes of the apparent motion was expected to captured by the sound conditions and the effect of sound was expected to be larger relative to the apparent motion with horizontal connecting bar. However expected trend was only observed to be consistent over the central channels.

In total, larger effect of sound for the vertical connecting bar condition over the frontal and temporal channels and for the single baseline condition over the central channels was unexpected finding. The reason for the trend observed over the frontal and temporal channels may be due to the attentional processes that multiple apparent motions led to since similar trend was not observed for the single apparent motion condition. Moreover larger difference between AV sound conditions for the vertical and single baseline conditions was observed around the later time intervals over the occipital-parietal channels. Nevertheless, we also observed larger individual effect of sound conditions over the vertical connecting bar condition around early time intervals of occipital-parietal channels. Further analyses on different early and later time intervals for all ROIs are needed to obtain more consistent results.

These exploratory findings obtained from the EEG data demonstrates that the effects of connecting bar conditions on audiovisual interaciton were present in low-level brain areas around early time intervals but further analyses which control the effect of attentional processes and examine the earlier time intervals in detail are needed to understand these neural mechanisms.

4.2 Spatial Influences on Cross-Modal Attention

The procedure multisensory studies on spatial attention often use is defining one modality as primary (task related) and the other modality as secondary one (task-irrelevant) by instructing participants to attend to the primary modality while ignoring the secondary [14, 54, 55]. Previous studies suggested that the exogenous shift of attention to the secondary modality led to its dominance over the primary modality [56] even though the primary modality was selectively attended [13]. This cross-modal attention paradigm is generated when stimuli from primary and secondary modalities share common properties [57]. Behavioral results of the current study seem to be in accordance with the cross-modal attention paradigms. When multiple apparent motions were presented with horizontal connecting bars between the flashes, participants had to distribute their attention to more than one component which made crossmodal attention between the auditory and visual systems harder. On the other hand, when these multiple apparent motions were grouped in one global apparent motion, participants focused their attention only to this bigger element which made the crossmodal attention more effective. In turn, this change in crossmodal attention resulted in larger effect of auditory stimulus on visual stimulus. Thus, spatial properties of primary modality (number of stimuli) had affected crossmodal attention.

Parietal and prefrontal regions were regarded as multimodal regions [58, 59]. Previous studies indicated the activation of bilateral inferior parietal and prefrontal regions in response to the temporal order judgment [60] and simultaneity taks [61] with audiovisual stimuli. Current study also indicated both early and later audiovisual patterns over the parietal and frontal channels for all connecting bar conditions. However these regions were also found to be activated within the frontoparietal attention network in response to tasks requiring selective attention. The frontoparietal attention network is the dorsal brain regions that are known to be activated in tasks requiring the shift of attention between different modalities and selectively attend to one modality over other. Namely, these regions are posterior parietal, premotor, prefrontal cortices (including frontal eye field) and parietal areas 7a and 7b in inferior parietal lobule [62]. Previous studies observed activity of inferior parietal and posterior prefrontal regions in response to both auditory and visual orienting tasks [63, 64], feature based attention [62]. Therefore the observed activations over the parietal and frontal channels may be the consequence of selective attention, not audiovisual interaction. Further experiments in which the effect of selective attention is controlled will be informative in this respect.

4.3 Future Directions

Mercier and colleagues [34] suggested that the supra-additive and sub-additive models for observed audiovisual effects are larger and smaller respectively relative to the summation of activities associated with the processing of uni-modal stimuli. Results of the current study could fit to the supra-additive or sub-additive models since the audiovisual P1s, N1s and P2s were found to be different than the visual-only condition. Further analysis comparing the summation of unimodal stimulus (auditory-only + visual-only) to the audiovisual conditions (AV) are needed. Moreover, the current experiment here investigated only the amplitude differences on averaged ERPs. Comparison of the latencies of the averaged ERPs may give an idea about the exact timing of P1, N1 and P2s across connecting bar conditions and extend the findings of the current study. Also, wavelet transform analysis can be applied in order to interpret the EEG data on time-frequency window.

Bibliography

[1] P. Bertelson and G. Aschersleben, "Automatic visual bias of perceived auditory location," *Psychonomic Bulletin & Review*, vol. 5, no. 3, pp. 482–489, 1998.

[2] I. P. Howard and W. B. Templeton, "Human spatial orientation.," 1966.

[3] R. Fendrich and P. M. Corballis, "The temporal cross-capture of audition and vision," *Perception & Psychophysics*, vol. 63, no. 4, pp. 719–725, 2001.

[4] S. Morein-Zamir, S. Soto-Faraco, and A. Kingstone, "Auditory capture of vision: examining temporal ventriloquism," *Cognitive Brain Research*, vol. 17, no. 1, pp. 154–163, 2003.

[5] G. H. Recanzone, "Auditory influences on visual temporal rate perception," *Journal of Neurophysiology*, vol. 89, no. 2, pp. 1078–1093, 2003.

[6] D. Alais and D. Burr, "No direction-specific bimodal facilitation for audiovisual motion detection," *Cognitive Brain Research*, vol. 19, no. 2, pp. 185–194, 2004.

[7] H. Kafaligonul and G. R. Stoner, "Auditory modulation of visual apparent motion with short spatial and temporal intervals," *Journal of Vision*, vol. 10, no. 12, pp. 31–31, 2010.

[8] S. Getzmann, "The effect of brief auditory stimuli on visual apparent motion," *Perception*, vol. 36, no. 7, pp. 1089–1103, 2007.

[9] S. Palmer and I. Rock, "Rethinking perceptual organization: The role of uniform connectedness," *Psychonomic Bulletin & Review*, vol. 1, no. 1, pp. 29–55, 1994.

[10] R. N. Shepard and S. L. Zare, "Path-guided apparent motion," *Science*, vol. 220, no. 4597, pp. 632–634, 1983.

[11] M. Radeau, "Auditory-visual spatial interaction and modularity.," *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 1994.

[12] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, pp. 746–748, 1976.

[13] L. Ortega, E. Guzman-Martinez, M. Grabowecky, and S. Suzuki, "Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability," *Attention, Perception, & Psychophysics*, vol. 76, no. 5, pp. 1485–1502, 2014.

[14] K. Yuasa and Y. Yotsumoto, "Opposite distortions in interval timing perception for visual and auditory stimuli with temporal modulations," *PloS One*, vol. 10, no. 8, p. e0135646, 2015.

[15] R. Sekuler, A. B. Sekuler, and R. Lau, "Sound alters visual motion perception.," *Nature*, vol. 385, no. 6614, p. 308, 1997.

[16] E. Freeman and J. Driver, "Direction of visual apparent motion driven solely by timing of a static sound," *Current Biology*, vol. 18, no. 16, pp. 1262–1266, 2008.

[17] T. Z. Strybel, C. L. Manligas, O. Chan, and D. R. Perrott, "A comparison of the effects of spatial separation on apparent motion in the auditory and visual modalities," *Perception & Psychophysics*, vol. 47, no. 5, pp. 439–448, 1990.

[18] L. Shams, Y. Kamitani, and S. Shimojo, "Illusions: What you see is what you hear," *Nature*, vol. 408, no. 6814, pp. 788–788, 2000.

[19] R. B. Welch and D. H. Warren, "Immediate perceptual response to intersensory discrepancy.," *Psychological Bulletin*, vol. 88, no. 3, p. 638, 1980.

[20] A. Vatakis and C. Spence, "Evaluating the influence of the ?unity assumption?on the temporal perception of realistic audiovisual stimuli," *Acta Psychologica*, vol. 127, no. 1, pp. 12–23, 2008.

[21] T. Shipley, "Auditory flutter-driving of visual flicker," *Science*, vol. 145, no. 3638, pp. 1328–1330, 1964.

[22] A. S. Bregman, "The formation of auditory streams," *Attention and Performance*, vol. 7, pp. 63–76, 1978.

[23] A. S. Bregman and A. Achim, "Visual stream segregation," *Perception & Psychophysics*, vol. 13, no. 3, pp. 451–454, 1973.

[24] M. Wertheimer, "A brief introduction to gestalt, identifying key theories and principles," *Psychol. Forsch*, vol. 4, pp. 301–350, 1923.

[25] L. Chen, "Topological perception? a possible dark cloud over computational theory," *Essays on Cognitive Sciences*, pp. 250–301, 1986.

[26] S. Han, G. W. Humphreys, and L. Chen, "Uniform connectedness and classical gestalt principles of perceptual grouping," *Perception & Psychophysics*, vol. 61, no. 4, pp. 661–674, 1999.

[27] D. D. Kurylo, "Time course of perceptual grouping," *Perception & Psychophysics*, vol. 59, no. 1, pp. 142–147, 1997.

[28] A. Ginsburg, "Pattern recognition techniques suggested from psychological correlates of a model of the human visual system," *NAECON 73*, pp. 309–316, 1973.

[29] S. Grossberg and E. Mingolla, "Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading.," *Psychological Review*, vol. 92, no. 2, p. 173, 1985.

[30] E. Schröger and A. Widmann, "Speeded responses to audiovisual signal changes result from bimodal integration," *Psychophysiology*, vol. 35, no. 6, pp. 755–759, 1998.

[31] J. Miller, "Channel interaction and the redundant-targets effect in bimodal divided attention.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 17, no. 1, p. 160, 1991.

[32] M. H. Giard and F. Peronnet, "Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study," *Journal of Cognitive Neuroscience*, vol. 11, no. 5, pp. 473–490, 1999.

[33] B. Bonath, T. Noesselt, A. Martinez, J. Mishra, K. Schwiecker, H.-J. Heinze, and S. A. Hillyard, "Neural basis of the ventriloquist illusion," *Current Biology*, vol. 17, no. 19, pp. 1697–1703, 2007.

[34] M. R. Mercier, S. Molholm, I. C. Fiebelkorn, J. S. Butler, T. H. Schwartz, and J. J. Foxe, "Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation," *The Journal of Neuroscience*, vol. 35, no. 22, pp. 8546–8557, 2015.

[35] Q. Li, Y. Wu, J. Yang, J. Wu, and T. Touge, "The temporal reliability of sound modulates visual detection: An event-related potential study," *Neuroscience Letters*, vol. 584, pp. 202–207, 2015.

[36] D. Bueti, B. Bahrami, and V. Walsh, "Sensory and association cortex in time perception," *Journal of Cognitive Neuroscience*, vol. 20, no. 6, pp. 1054–1062, 2008.

[37] M. Azuma and H. Suzuki, "Properties and distribution of auditory neurons in the dorsolateral prefrontal cortex of the alert monkey," *Brain Research*, vol. 298, no. 2, pp. 343–346, 1984.

[38] L. A. Benevento, J. Fallon, B. Davis, and M. Rezak, "Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey," *Experimental Neurology*, vol. 57, no. 3, pp. 849–872, 1977.

[39] B. E. Stein and M. A. Meredith, *The merging of the senses.* The MIT Press, 1993.

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

[41] D. Sanabria, S. Soto-Faraco, J. S. Chan, and C. Spence, "When does visual perceptual grouping affect multisensory integration?," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 4, no. 2, pp. 218–229, 2004.

[42] M. Keetels and J. Vroomen, "Sound affects the speed of visual processing.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 3, p. 699, 2011.

[43] F. A. Wichmann and N. J. Hill, "The psychometric function: I. fitting, sampling, and goodness of fit," *Perception & Psychophysics*, vol. 63, no. 8, pp. 1293–1313, 2001.

[44] A. King and A. Palmer, "Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus," *Experimental Brain Research*, vol. 60, no. 3, pp. 492–500, 1985.

[45] H. Ohmura, "Intersensory influences on the perception of apparent movement," *Japanese Psychological Research*, vol. 29, no. 1, pp. 1–9, 1987.

[46] P. S. Goldman-Rakic, "Topography of cognition: parallel distributed networks in primate association cortex," *Annual Review of Neuroscience*, vol. 11, no. 1, pp. 137–156, 1988.

[47] M. M. Murray, A. Thelen, G. Thut, V. Romei, R. Martuzzi, and P. J. Matusz, "The multisensory function of the human primary visual cortex," *Neuropsychologia*, vol. 83, pp. 161–169, 2016.

[48] W. Teder-Sälejärvi, J. McDonald, F. Di Russo, and S. Hillyard, "An analysis of audio-visual crossmodal integration by means of event-related potential (erp) recordings," *Cognitive Brain Research*, vol. 14, no. 1, pp. 106–114, 2002.

[49] W. Yang, Q. Li, T. Ochi, J. Yang, Y. Gao, X. Tang, S. Takahashi, and J. Wu, "Effects of auditory stimuli in the horizontal plane on audiovisual integration: an event-related potential study," *PloS One*, vol. 8, no. 6, p. e66402, 2013.

[50] M. G. Woldorff, C. C. Gallen, S. A. Hampson, S. A. Hillyard, C. Pantev, D. Sobel, and F. E. Bloom, "Modulation of early sensory processing in human auditory cortex during auditory selective attention," *Proceedings of the National Academy of Sciences*, vol. 90, no. 18, pp. 8722–8726, 1993.

[51] D. Talsma and M. G. Woldorff, "Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity," *Journal of Cognitive Neuroscience*, vol. 17, no. 7, pp. 1098–1114, 2005.

[52] S. Yuval-Greenberg and L. Y. Deouell, "What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition," *The Journal of Neuroscience*, vol. 27, no. 5, pp. 1090–1096, 2007.

[53] R. Cecere, G. Rees, and V. Romei, "Individual differences in alpha frequency drive crossmodal illusory perception," *Current Biology*, vol. 25, no. 2, pp. 231–235, 2015.

[54] M. Eimer and E. Schröger, "Erp effects of intermodal attention and cross-modal links in spatial attention," *Psychophysiology*, vol. 35, no. 03, pp. 313–327, 1998.

[55] W. A. Teder-Sälejärvi, T. F. Münte, F.-J. Sperlich, and S. A. Hillyard, "Intra-modal and cross-modal spatial attention to auditory and visual stimuli. an event-related brain potential study," *Cognitive Brain Research*, vol. 8, no. 3, pp. 327–343, 1999.

[56] J. Driver and C. Spence, "Cross–modal links in spatial attention," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 353, no. 1373, pp. 1319–1331, 1998.

[57] V. M. Ciaramitaro, G. T. Buračas, and G. M. Boynton, "Spatial and cross-modal attention alter responses to unattended sensory information in early

visual and auditory human cortex," *Journal of Neurophysiology*, vol. 98, no. 4, pp. 2399–2413, 2007.

[58] C.-T. Wu, D. Weissman, K. Roberts, and M. Woldorff, "The neural circuitry underlying the executive control of auditory spatial attention," *Brain Research*, vol. 1134, pp. 187–198, 2007.

[59] A. Degerman, T. Rinne, A.-K. Särkkä, J. Salmi, and K. Alho, "Selective attention to sound location or pitch studied with event-related brain potentials and magnetic fields," *European Journal of Neuroscience*, vol. 27, no. 12, pp. 3329–3341, 2008.

[60] M. Binder, "Neural correlates of audiovisual temporal processing–comparison of temporal order and simultaneity judgments," *Neuroscience*, vol. 300, pp. 432–447, 2015.

[61] R. Ptak, "Neural correlates of auditory-visual stimulus onset ansynchrony detection," *Journal of Neuroscience*, vol. 21, no. 1, pp. 300–304, 2001.

[62] R. Ptak, "The frontoparietal attention network of the human brain action, saliency, and a priority map of the environment," *The Neuroscientist*, vol. 18, no. 5, pp. 502–515, 2012.

[63] J. Salmi, T. Rinne, A. Degerman, O. Salonen, and K. Alho, "Orienting and maintenance of spatial attention in audition and vision: multimodal and modality-specific brain activations," *Brain Structure and Function*, vol. 212, no. 2, pp. 181–194, 2007.

[64] M. Corbetta, G. Patel, and G. L. Shulman, "The reorienting system of the human brain: from environment to theory of mind," *Neuron*, vol. 58, no. 3, pp. 306–324, 2008.

Appendix A

Paired-samples t-tests across 63 channels

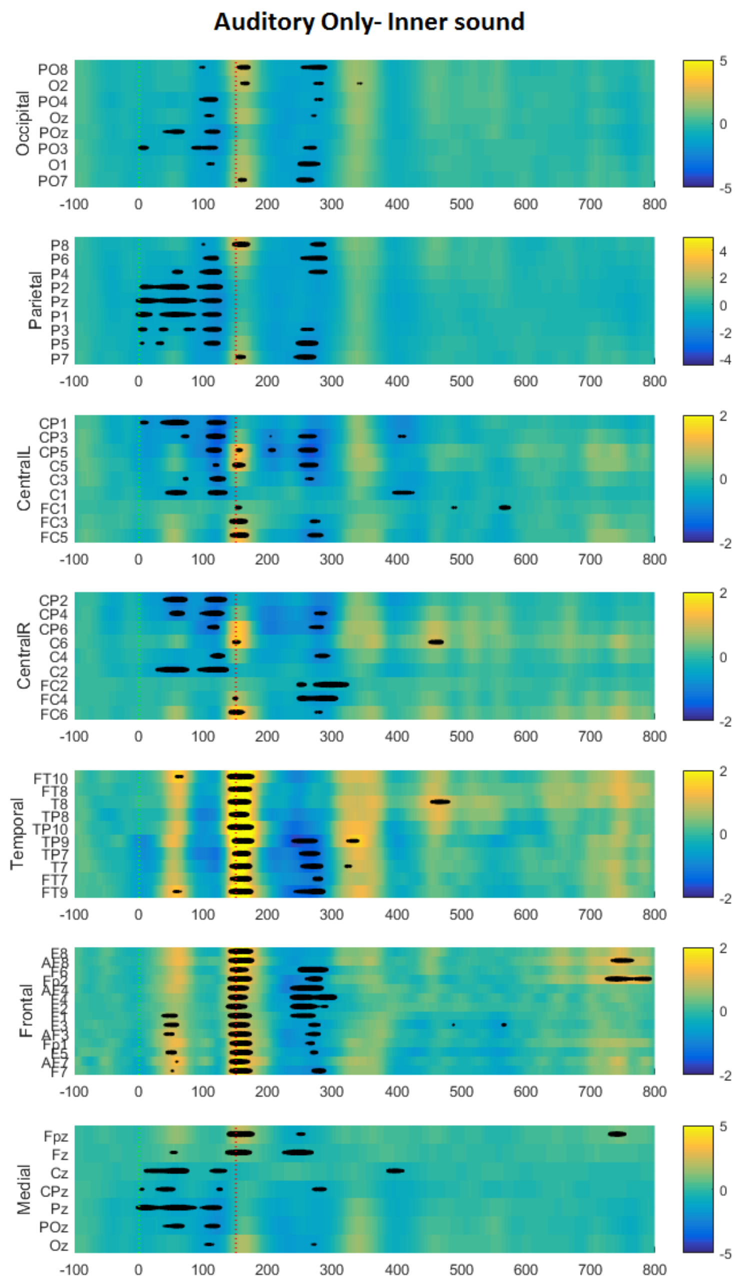

Figure A.1: Paired-samples t-test on 63 channels to assess whether auditory-only inner sound condition is significantly different than baseline

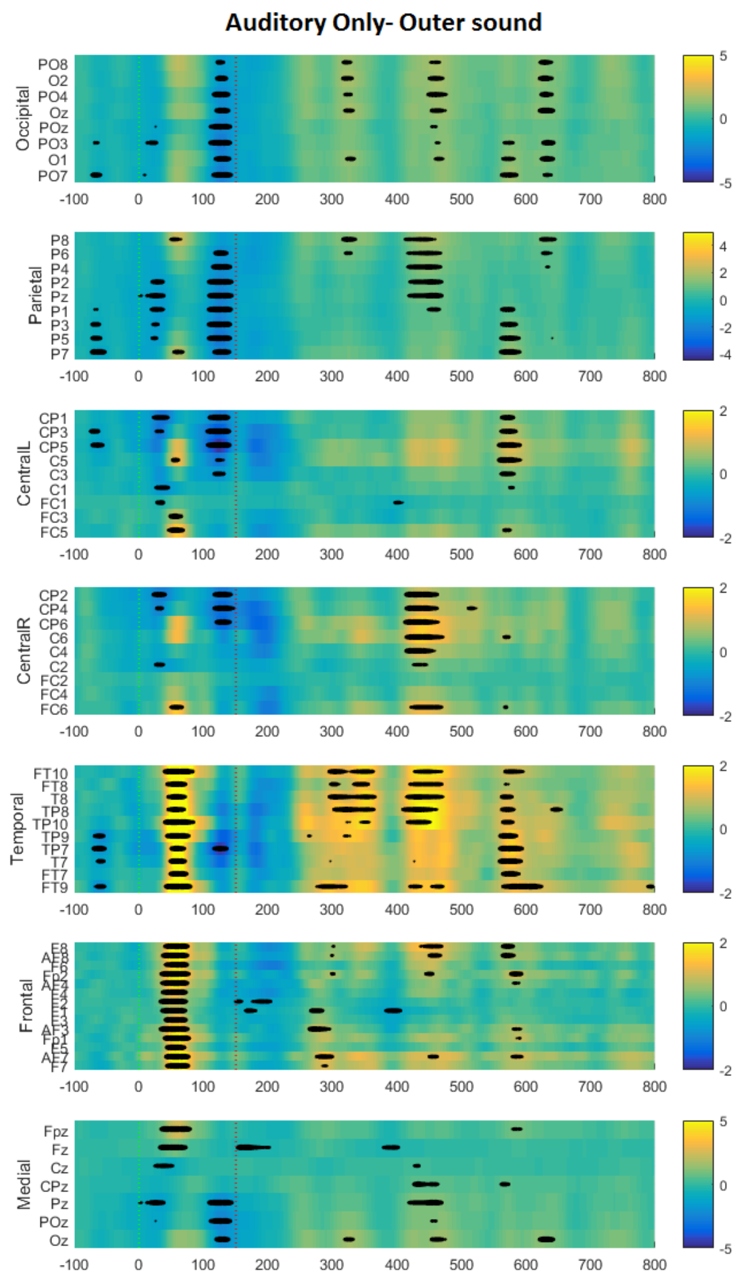

Figure A.2: Paired-samples t-test on 63 channels to assess whether auditory-only outer sound condition is significantly different than baseline

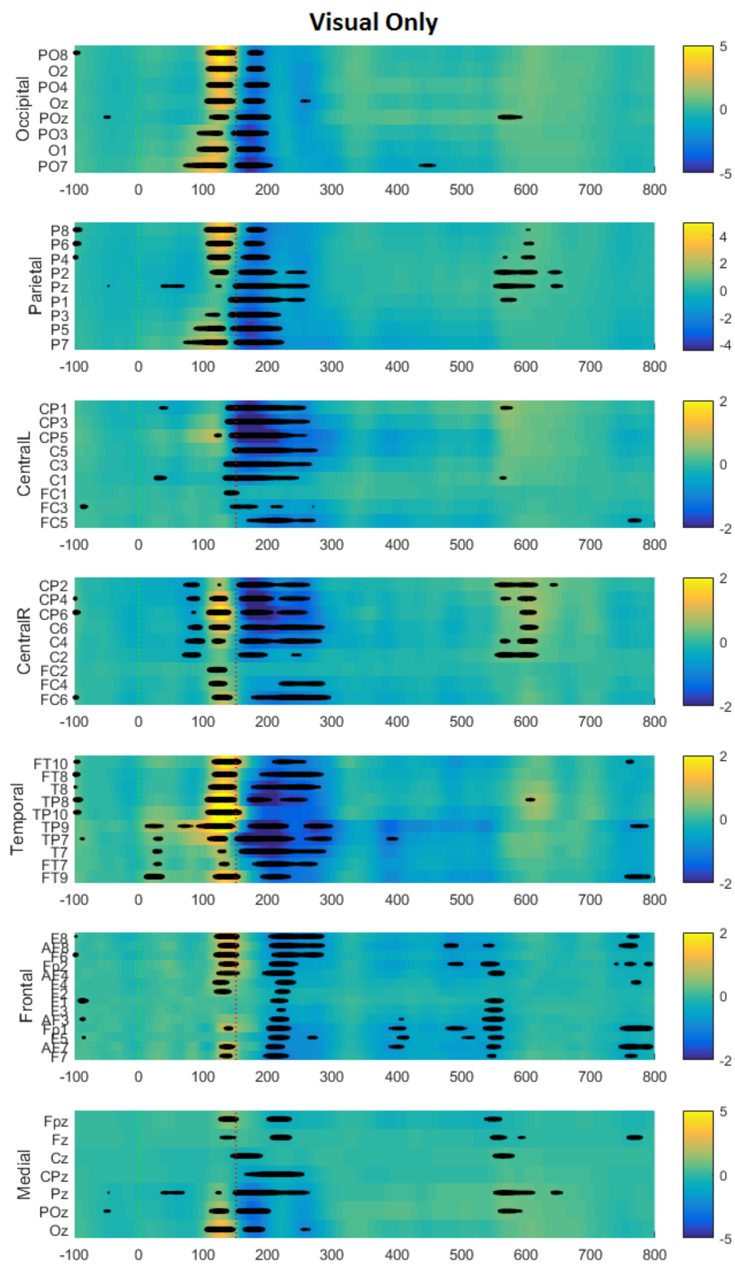

Figure A.3: Paired-samples t-test on 63 channels to assess whether visual-only condition is significantly different than baseline

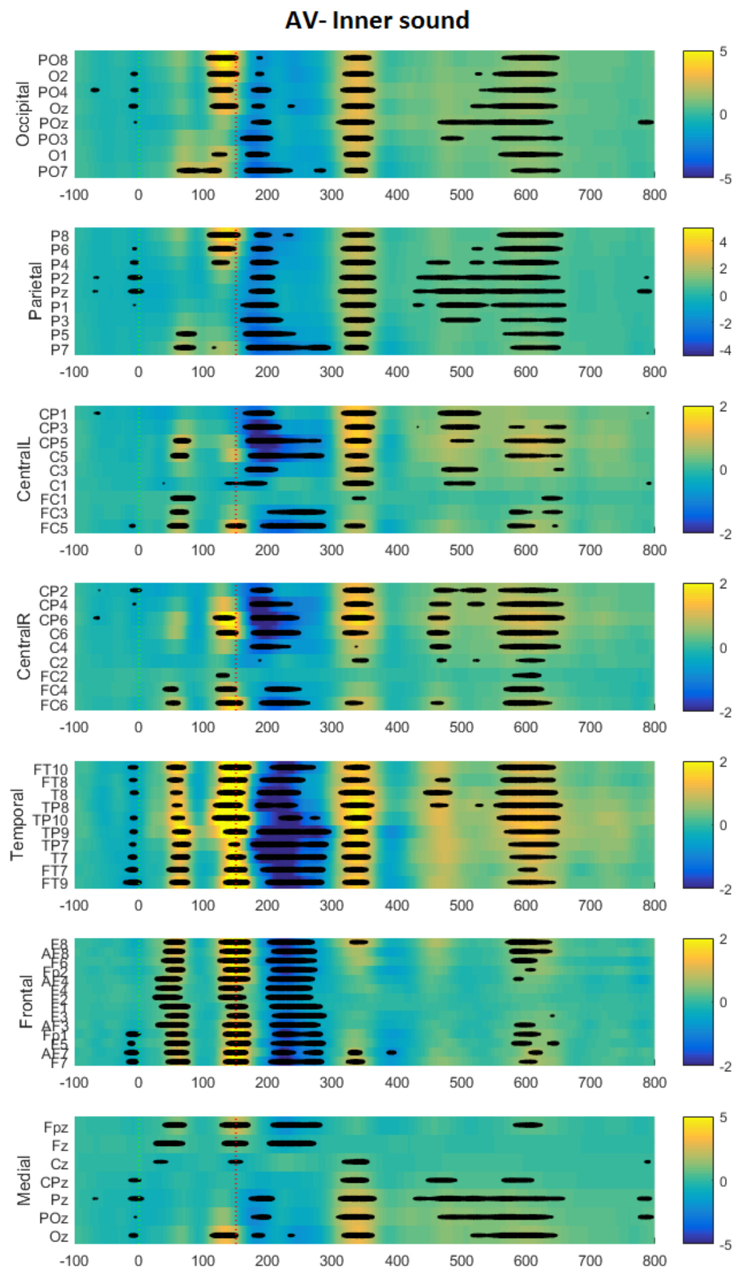

Figure A.4: Paired-samples t-test on 63 channels to assess whether AV inner sound condition is significantly different than baseline

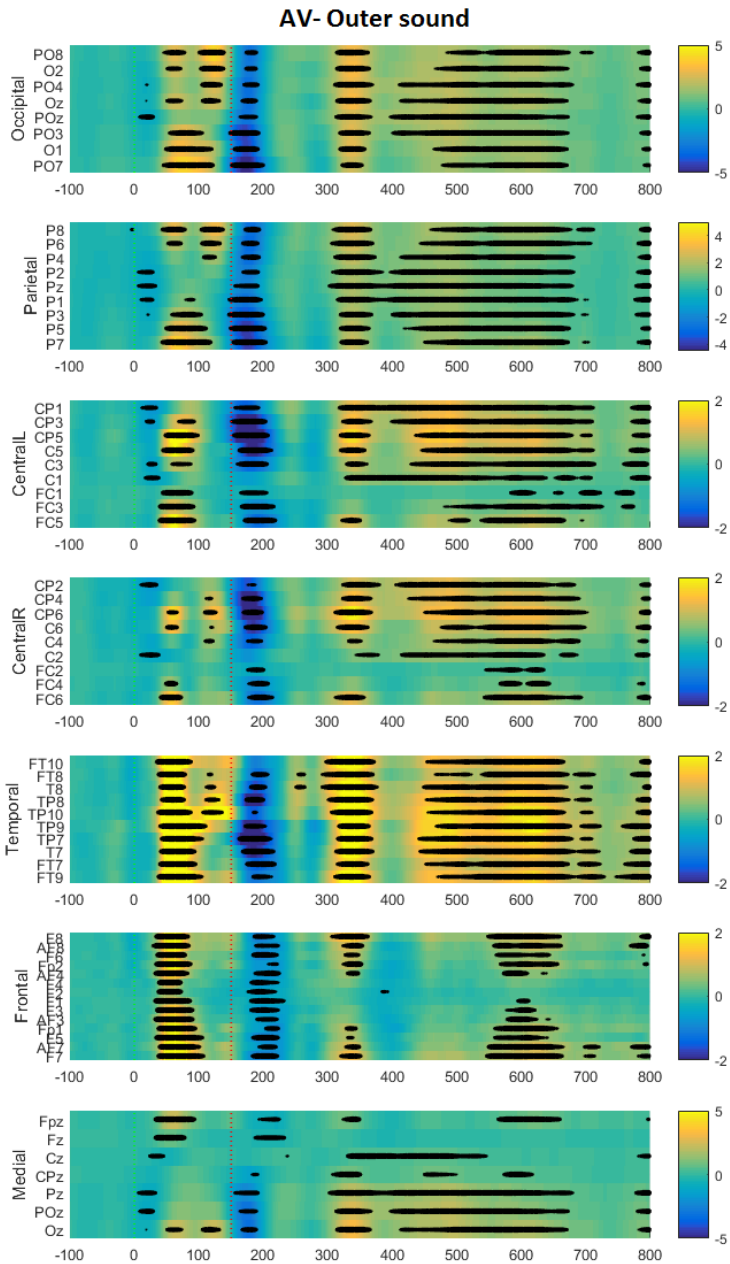

Figure A.5: Paired-samples t-test on 63 channels to assess whether AV outer sound condition is significantly different than baseline

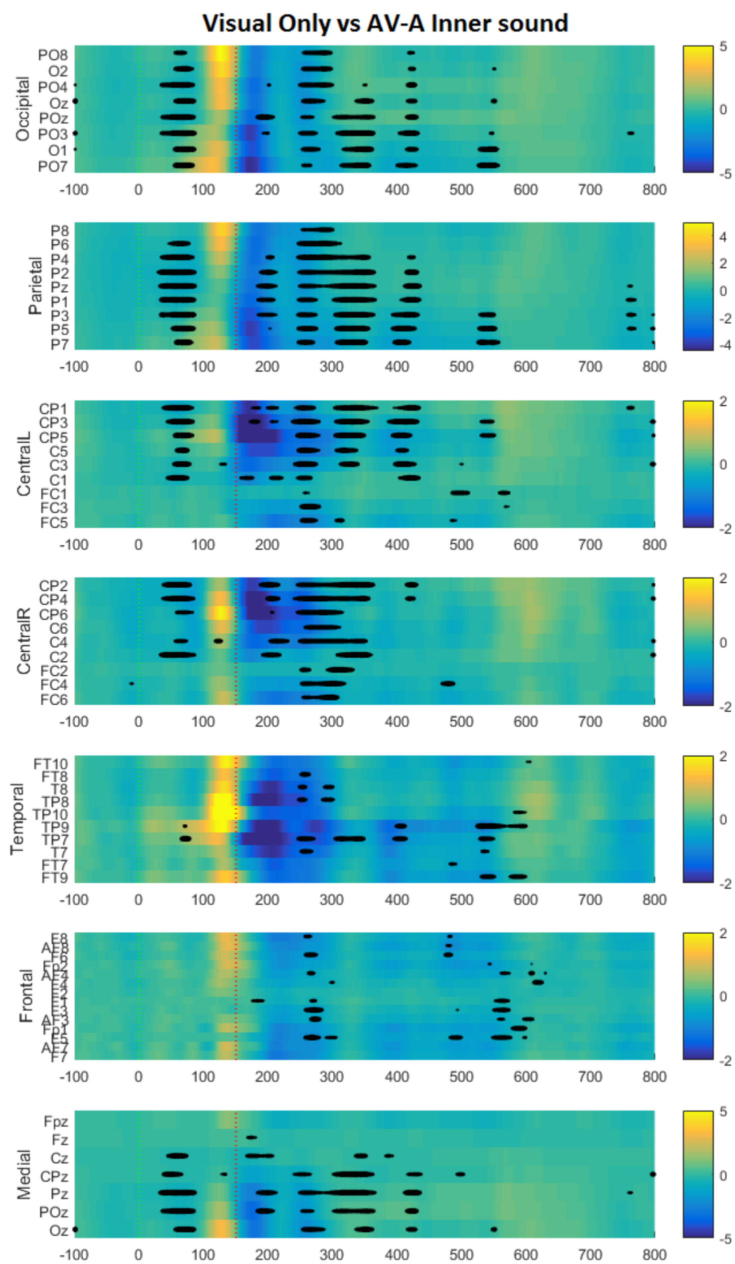

Figure A.6: Paired-samples t-test on 63 channels to assess whether AV-A inner sound condition is significantly different than visual-only condition

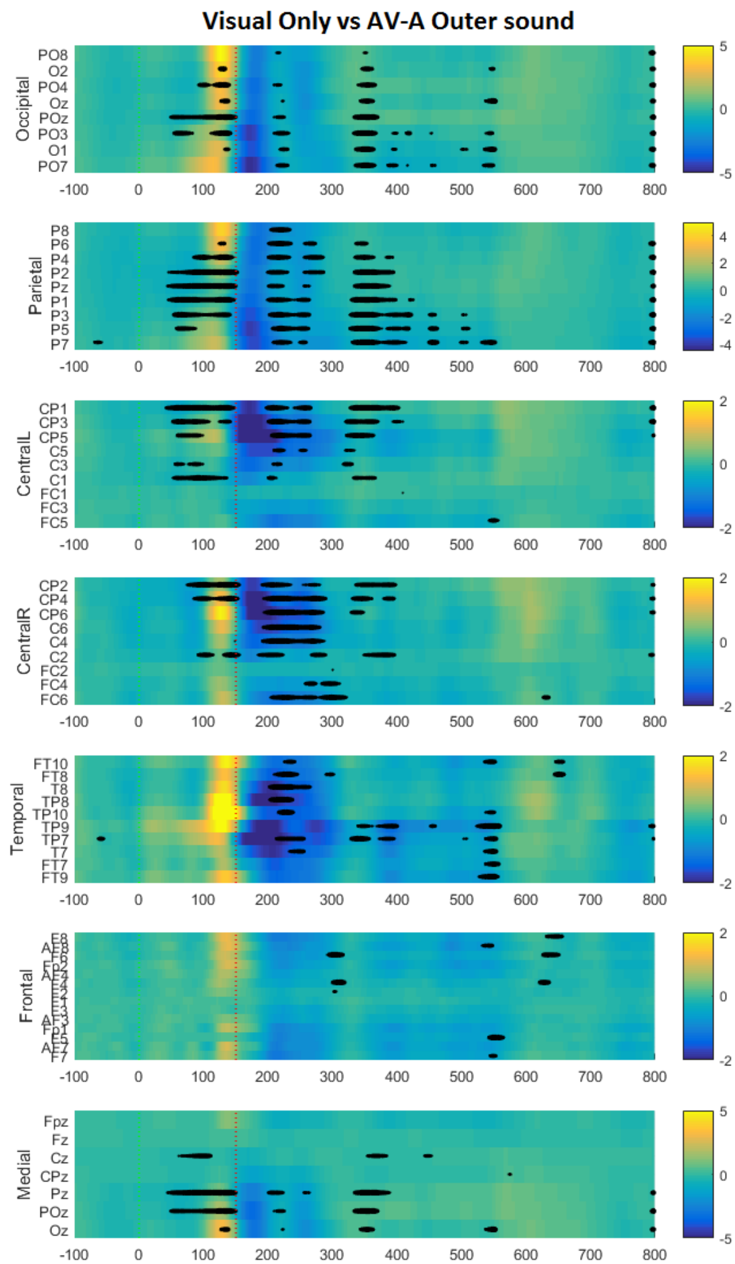

Figure A.7: Paired-samples t-test on 63 channels to assess whether AV-A outer sound condition is significantly different than visual-only condition

Appendix B

Descriptive Statistics Tables

ConnectingBars * Proximity

ConnectingBars	Proximity	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Horizontal	1.5	71,684	5,671	59,432	83,936
	2.7	65,561	4,229	56,425	74,698
	4.4	68,878	4,888	58,319	79,436
Vertical	1.5	77,551	4,815	67,148	87,954
	2.7	78,827	4,243	69,659	87,994
	4.4	81,378	3,476	73,869	88,886
2 Aligned	1.5	78,571	4,350	69,174	87,969
	2.7	82,143	3,781	73,974	90,311
	4.4	75,510	5,076	64,543	86,477

Table B.1: Mean values, standard errors and standard deviations of connecting bar conditions for each proximity condition

Proximity * CommonFatePrinciple

Proximity	CommonFatePrinciple	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.5	Common fate	78,571	4,350	69,174	87,969
	2 Aligned	81,888	3,252	74,862	88,914
2.7	Common fate	82,143	3,781	73,974	90,311
	2 Aligned	76,276	4,745	66,025	86,526
4.4	Common fate	75,510	5,076	64,543	86,477
	2 Aligned	73,469	4,642	63,442	83,497

Table B.2: Mean values, standard errors and standard deviations of common fate conditions for each proximity condition

ConnectingBars * sound

ConnectingBars	sound	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Horizontal	Inner	104,640	10,623	77,333	131,947
	Outer	138,667	9,266	114,848	162,486
Vertical	Inner	91,315	8,256	70,091	112,539
	Outer	147,167	7,669	127,452	166,881

Table B.3: Mean values, standard errors and standard deviations of connecting bar conditions for each sound condition

ConnectingBars * sound

ConnectingBars	sound	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Single Horizontal	Inner	95,575	11,178	69,143	122,007
	Outer	135,970	9,646	113,161	158,779
Single Baseline	Inner	84,713	12,708	54,664	114,761
	Outer	166,788	13,860	134,013	199,562

Table B.4: Mean values, standard errors and standard deviations of single apparent motion condition with different connecting bar conditions for each sound condition