Fall 2023

Neural Correlates of Audiovisual Integration

Jonathan Conrady

Recommended Citation

Conrady, J.(2023). *Neural Correlates of Audiovisual Integration.* (Master's thesis). Retrieved from https://scholarcommons.sc.edu/etd/7602

Neural Correlates of Audiovisual Integration

by

Jonathan Conrady

Associate of Arts
Greenville Technical College, 2019

Bachelor of Science
University of South Carolina Upstate, 2021

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Arts in

Experimental Psychology

College of Arts and Sciences

University of South Carolina

2023

Accepted by:

Jessica Green, Director of Thesis

Svetlana Shinkareva, Reader

Caitlin Hudac, Reader

Ann Vail, Dean of the Graduate School

ABSTRACT

Successfully integrating multiple streams of information is paired with a variety of behavioral enhancements. Multisensory research over the past few decades has demonstrated the criticality of spatial proximity and stimulus effectiveness. Here, we focused on the temporal aspects and underlying neural activity of frequency-varying, audiovisual stimuli using scalp recorded electroencephalography (EEG). Our set of stimuli was particularly chosen because humans are most sensitive to the auditory sound and visual spatial frequencies that make up speech. Recent work has shown that audiovisual integration effects of low-level, simple stimuli are largest when both modalities fall within, compared to outside of, their respective sensitivity ranges. At the electrophysiological level, we found evidence for bilateral visual modality effects at each SOA condition, which resulted in speeded visual N1 latency and enhanced P2 amplitude event related potentials (ERPs). Additionally, when the stimuli were separated in time, we observed differences in sensory ERPs for stimuli that were perceived as simultaneous compared to those that were not integrated, as well as variations in the ERP effects based on the frequency range of the stimuli. These findings offer insight at the neural level that human sensitivity to low-level stimulus features likely plays an important role in the enhanced integration for audiovisual speech.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The rapid integration of multiple sensory inputs is necessary for humans to accurately interpret the vast amount of information presented in the environment. Literature in the field of multisensory integration has consistently shown that the processing of multimodal stimuli, compared to unimodal stimuli, is typically enhanced as demonstrated in speeded response times (RTs) and increased detection and localization capabilities (Donohue, Green, & Woldorff, 2015; Giard & Perronet, 1999; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Murray et al., 2016; Nelson et al., 1998; Stein & Standford, 2008; Stevenson & Wallace, 2013). Studies investigating subcortical structures of nonhuman animals (e.g., cats, rats, mice, and monkeys) have discovered that the superior colliculus is a prominent structure linked to the integration of multiple stimuli, as it contains neurons that emit signals in response to unimodal stimuli and others that emit significantly greater signals in response to multimodal stimuli due to the overlap of receptive fields (Giard & Perronet, 1999; Wallace, Meredith, & Stein, 1998). The outcome of these findings led to the emergence of three key properties involved in integration effects: stimulus effectiveness, spatial proximity, and temporal proximity (Spence, 2013). Stimulus effectiveness refers to how effective the stimulus is at eliciting a neural or behavioral response. Unisensory inputs that are only weakly effective are more readily integrated compared to highly salient ones (i.e., the rule of inverse effectiveness; Meredith & Stein, 1986; Stein & Standford 2008). This is supported by behavioral evidence in which poor performance during unisensory conditions is later reflected by the strongest display of multisensory integration effects in the same group of participants (Holmes, 2007).

Successful integration of multiple streams of sensory information is also linked with spatial and temporal constraints; that is, information from different modalities must occur relatively close together in space and time for a resulting bimodal percept (Chen and Spence, 2017; Vatakis and Spence, 2007; Koelewijn, Bronkhorst, & Theeuwes, 2010). The implications of spatial correspondence during crossmodal binding have been supported by covert and overt attention paradigms, both of which have demonstrated that detection of a target in one modality is facilitated when an accessory stimulus of a different modality accompanies the target in the same, rather than different, hemifield (Spence, 2013). This suggests that our spatial integration capabilities are highly sensitive, largely due to the nature of our receptive fields in the superior colliculus (Chen, 2013; Wallace et al., 1998).

Along with spatial factors, it is crucial to consider the temporal aspects of multisensory integration, which can govern whether two unisensory inputs are streamed together into a single bimodal percept. The temporal properties of integration are best represented as a negative linear trend, in that crossmodal effects are highest the closer two stimuli onset in time. More specifically, integration is strongest when the stimuli onset less than 100 milliseconds apart (Koelewijn et al., 2010). This was supported by Meredith, Nemitz, and Stein (1987), who demonstrated that multisensory neurons in cats diminish in response as the two unisensory stimuli dissipate further in time. However, temporal integration effects have been reported even up to a few hundred milliseconds of onset differences, as demonstrated with the processing of audiovisual speech in humans (Spence & Squire, 2003; Stevenson & Wallace, 2013; van Wassenhove, Grant & Poeppel, 2005; Vatakis & Spence, 2010). It has been established that a critical time period exists in which integration can occur despite temporal discrepancies of stimulus onset (Dixon & Spitz, 1980; Stevenson and Wallace, 2013); this has been classified as the temporal window of integration (TWI). Although two stimuli may occur at differing time periods, if the discrepancy is within the TWI, our brain will not perceive them as separable occurrences and instead interpret the two stimuli as a concurrent event (Vroomen & Keetels, 2010; Donohue et al., 2015). Increasing the width of the TWI is associated with increased integration, despite greater temporal disparity between the stimuli, which may or may not be beneficial depending on the circumstances.

Audiovisual integration in general, including TWI width, appears to be modulated by a variety of spatiotemporal and semantic factors (Chen & Spence, 2017; Spence & Squire, 2003; Vatakis & Spence, 2007), attention and task instructions (Donohue et al., 2015), modality order (Cecere et al., 2017; Stevenson & Wallace, 2013), and stimulus intensity and complexity (Green et al., 2019; Stevenson & Wallace, 2013). For TWI specifically, stimulus complexity has been shown to modulate TWI width, with human audiovisual speech being integrated at much larger temporal discrepancies than simple stimuli (Dixon & Spitz, 1980; Stevenson & Wallace, 2013). However, the TWI width also appears to be modulated by low-level features of the unimodal stimuli. Green and colleagues (2019) established that the low-level features most common to speech stimuli result in TWI width increases compared to those low-level features that are not typical in speech. However, the neural signatures associated with the integration of low-level, audiovisual features and how they relate to TWI width are less well understood. The proposed research objective is to examine the factors influencing multisensory integration and TWI modulation, and to offer novel evidence of the underlying event-related brain potential (ERP) components present in the integration of low-level, audiovisual stimuli.

1.1 FACTORS INFLUENCING MULTISENSORY INTEGRATION AND TWI WIDTH

Visual and auditory information can influence each other at the semantic, spatial, and temporal level to modulate multisensory integration and TWI width (Chen & Spence, 2017; Spence & Squire, 2003). The unity effect refers to stimuli being more readily integrated when they are perceived as sharing the same event or object (Chen and Spence, 2017; Vatakis and Spence, 2007). This effect is typically experienced unconsciously, as is the case with observing a barking dog. The semantically congruent occurrence of the dog's mouth movement and barking sounds allow for strong integration effects and minimize any binding uncertainty as these events originate from the same spatiotemporal source (i.e., the dog) and should go together (Chen & Spence, 2017).

Similar processes occur during spatial and temporal ventriloquism effects, in which the brain attempts to correct for spatial and temporal asynchronies, respectively, through modality-specific indications (Spence & Squire, 2003). During spatial ventriloquism, auditory stimuli are synchronized towards the visual spatial event (Koelewijn et al., 2010; Spence & Squire, 2003), resulting in bimodal perception. An example of this is watching a movie in a theater. Despite the projection of the actors' voices through speakers at different locations from the screen, audience members will perceive the origin of the sounds as being generated from the actors' mouths, rather than identifiably from the speakers (Calvert, Spence, & Stein, 2004; Chen and Spence, 2017). Conversely, the temporal ventriloquism effect demonstrates how the temporal binding between visual and auditory information can correct for an asynchrony, resulting in a bimodal percept. For example, the length perception of a light flashing at unvaried rates is modulated by the rate of accompanied sounds (Spence & Squire, 2003). Similarly, illusory multisensory events show that a static flash of light can be perceived as flickering when accompanied by repeated, short-termed bursts of tones (Koelewijn et al., 2010). Although these findings suggest that auditory and visual events can modulate each other, the integration of multiple stimuli (e.g., vision and audition) can result in complementation effects in which one modality benefits from the addition of the other in modality-specific tasks. Such complementation has been shown during simple audiovisual events. During a visual search task, the addition of a salient auditory stimulus can facilitate RT and target detection accuracy. This can produce what is known as the "freeze effect," due to subjects reporting that the visual stimulus appears to remain on the screen for a longer period of time when coupled with a sound (Koelewijn et al., 2010). Additionally, the presentation of short tones paired with a color-changing target has shown increased target detection performance. Complementation effects are also present in speech, as visual lip movements can offer predictive information of proceeding vocal sounds (Cecere, Gross, & Thut, 2016; Stekelenburg & Vroomen, 2007; Wassenhove et al., 2005).

Moreover, factors such as attention and task instructions can serve as significant modulators in integration processing. For example, Donohue et al. (2015) explored the avenue of potential factors that may influence TWI width and proposed that spatial attention capabilities likely play an important role. Utilizing the “bounce/stream paradigm” and auditory bounce effect (ABE), the researchers examined whether hemifield-specific cued spatial attention could modulate TWI width across discriminatory or simultaneity judgements (SJs) and stimulus onset asynchronies (SOAs; D0/D150/D300). Attention’s interaction with TWI were two-fold. First, when participants were instructed to discriminate between either a “bounce” or “stream” event as two moving circles overlapped, the TWI was observed to be much narrower. Second, during a task which required SJs between the circles’ visual overlap and a collision-like sound, participants’ TWI was recorded to be much broader. These results strongly indicate that spatial attention has a significant effect on the temporal properties of crossmodal binding and is likely task dependent (Donohue et al., 2015).

Audiovisual discrimination, as well as SJ tasks (Cecere et al., 2016; Green et al., 2019; Stevenson and Wallace, 2013), serve as measures to obtain information about TWI width by estimating SOAs at which individuals perceive two unisensory inputs as bimodal events (Donohue et al., 2015). Indeed, SJ tasks are commonly used in multisensory-related experiments and entail the simple judgement of whether two stimuli occurred at the same, or at different, times. By instructing participants to engage in SJ tasks, researchers are capable of measuring TWI width through varying increases of SOAs (Vroomen & Keetels 2010). Although an SOA at 0 ms reflects physically simultaneous events, the largest integration effects occur at the point of subjective simultaneity (PSS), or at which individuals perceive simultaneity. This typically occurs when visual information onsets prior to auditory information, as this occurrence is most natural in the world such as hearing thunder after seeing lightning (Spence & Squire, 2003; Stevenson & Wallace, 2013).

1.2 MODALITY DOMINANCE AND ORDER

The perception of audiovisual events is composed of the properties that make up their unisensory components. Each sensory system is specialized in tasks that are simply most appropriate for them (Modality Appropriateness Hypothesis; Calvert et al., 2004); for example, research has demonstrated the practicality of the visual system in experiments emphasizing spatial information (Calvert et al., 2004; Repp & Penel, 2002; Walker & Scott, 1981). This is particularly prominent in cases where the visual and auditory components of the same source are close together in spatial proximity and appear to be semantically congruent (spatial ventriloquism effect; Chen and Spence, 2017). Conversely, the auditory system has an advantage in temporal-related perception and tasks. For example, independently presenting a flash of light for one second and a sound for one second will result in an interpretation of the sound occurring for a longer period of time, despite the two stimuli occurring at physically identical lengths (Walker & Scott, 1981; Wearden et al., 1998). Furthermore, in determining changes of sequence regularity, auditory stimuli are more readily identified in comparison to visual stimuli (Repp & Penel, 2002). This suggests that modality dominance particularly during physically synchronous, audiovisual events is likely task dependent; however, perception and crossmodal integration may be modulated by modality order.

During an audiovisual event in which the visual input precedes the auditory input, individuals will perceive them as a simultaneous occurrence significantly more often in comparison to the reverse order (Spence & Squire, 2003; Stevenson & Wallace, 2013; Vatakis & Spence, 2010). The likely reason for this can be attributed to their physical behavior in space and our experience with lights and sounds. We constantly experience the differences of lights and sounds as they travel through space, in that, light waves reach the retina faster than sound waves reach the cochlea (Murray et al., 2016; Stevenson & Wallace, 2013; Spence & Squire, 2003). This further indicates the significance of temporal order and complementation of audiovisual stimuli during integration processing. Although Stevenson and Wallace (2013) used other stimuli (e.g., beeps, flashes, and tools) along with speech stimuli (e.g., syllabi utterances), it is likely that this discovered result of visual-then-auditory events resulting in stronger integration effects can be explained by our fixation on lip movements predicting concurrent vocal sounds (Cecere et al., 2016; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007). Thus, supporting the argument that speech processing is an influential factor in the binding process of multiple unisensory stimuli and the temporal order in which unisensory stimuli are presented can influence multisensory integration. In fact, it has been demonstrated at the behavioral level that integration properties such as TWI width are not symmetric across visual-leading and auditory-leading multisensory events. As a result, it has been argued that the integration process of multiple sensory inputs is not static; rather, processing of audiovisual stimuli recruits independent mechanisms depending on which modality leads (Cecere et al., 2016).

1.3 ELECTROPHYSIOLOGICAL EVIDENCE OF AUDIOVISUAL INTEGRATION

Over recent decades, the millisecond-level temporal precision of electroencephalography (EEG) and ERP have led to significant contributions regarding the temporal processes involved in crossmodal binding that could otherwise not have been established using neuroimaging techniques (van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007). Recent research conducted by Cecere and colleagues (2017) examined this topic to observe whether a profound difference in multisensory processing at the neural level exists between auditory-first and visual-first, bimodal events. Participants were instructed to complete SJ tasks while ERPs were recorded using a 128-channel EEG. With four conditions (i.e., A-only, V-only, AV, and VA,), the researchers conducted an additive model to isolate AV and VA activity [i.e., AV = AV - (A + V); VA = VA - (V + A)] and displayed difference activation across scalp topographic maps. Separate cortical activation patterns for VA (i.e., symmetric organization) and AV (i.e., right lateralized) events were reported. Consistent across both audiovisual stimulus conditions was activity patterns present at three stages in time: 38-95 ms, 142-222 ms, and 297-351 ms. These results suggest that audiovisual binding is a multistage process occurring as early as 40 ms (Giard & Perronet, 1999) in posterior-parietal regions. This trajectory of activation differences continued to occur in two subsequent stages located in central (142-222 ms) and anterior regions (297-351 ms; Cecere et al., 2017).

Stekelenburg and Vroomen (2007) compared ERPs between spoken phonemes and natural non-speech sounds during passive observation tasks to gain a better understanding of the auditory N1/P2 components, which have previously shown to be attenuated during audiovisual speech stimuli (van Wassenhove et al., 2005). Because speech is comprised of visual information (i.e., lip movement) preceding auditory information (i.e., vocal sound produced by the speaker), the researchers chose nonspeech stimuli with the same properties (e.g., spoon tapping against a mug and hands clapping). Participants observed four possible conditions of audiovisual stimuli: audiovisual synchrony, visual-only, auditory-only (with a blank screen for display), and a control (blank square without visual or sound information). The results showed attenuation in amplitude and latency of the auditory-evoked N1 and P2 components, as a representation of perceiving speech sounds accompanied by concurrent lip movements (van Wassenhove et al., 2005). However, it is worth noting that similar outcomes were found for natural nonspeech stimuli, suggesting that underlying neural mechanisms involved in bimodal processing of audiovisual stimuli may not be entirely restricted to speech (Stekelenburg and Vroomen, 2007). It is possible that this outcome could be attributed to the low-level features that compose speech and non-speech, bimodal events.

1.4 THE CURRENT STUDY

The current study aims to further explore the “specialness” of low-level features associated with speech in multisensory integration and to gain insight on the neural factors that lead to this particular type of audiovisual stimulus consistently resulting in the broadest TWI. Indeed, as humans who engage in verbal conversation on a daily basis, it has become evident that speech stimuli affect the temporal domain of crossmodal binding. For example, Stevenson and Wallace (2013) demonstrated this effect by utilizing a variety of tasks (i.e., SJ, temporal order judgment, perceptual fusion task, and a two-interval, forced-choice task) and stimuli (i.e., simple, complex nonspeech, and speech). The TWI showed increases in width and symmetry during events of speech, compared to simple and complex, nonspeech stimuli (Stevenson and Wallace, 2013: Wallace & Stevenson, 2014).

Although studies with complex speech and non-speech stimuli have suggested that speech may be “special” with regard to how it is integrated in the brain (Vatakis & Spence, 2010), recent work suggests that at least some of the differences observed for integration of speech stimuli may originate from differences in the low-level features of speech compared to non-speech stimuli. In particular, the visual spatial frequencies and auditory sound frequencies presented in lips and vocal speech, respectively, fall within the peak sensitivity range for human perception (Green et al., 2019). These peak frequencies that we are most sensitive to range from about 1000 to 4000 cycles per second (Hz) for sounds (Robinson and Dadson, 1956), and roughly 3 to 6 cycles per degree (cpd) for spatial information (De Valois et al.,1987). Green et al. (2019) observed a wider TWI for pairs of simple audio-visual stimuli that were both within the peak sensitivity range, similar to the widening of the TWI that has been observed for speech stimuli by Stevenson and Wallace (2013).

The most likely explanation for the “specialness” of speech is due to the natural fine-tuning trajectory of audiovisual speech integration capabilities that occur over the lifespan (Wallace & Stevenson, 2014) and can be segmented into the following three stages of multisensory integration: immature, broadly-tuned, and narrowly-tuned (Murray et al., 2016). As early as infancy, humans are capable of binding speech features such as the faces of caregivers and the vocalizations produced by them. After 8 months, this process becomes more fine-tuned with infants demonstrating selective attention to the speaker's mouth that is producing the source of speech sounds (Murray et al., 2016). As this developmental trajectory continues to childhood (i.e., around 10 years of age), the TWI still appears immature (Hillock, Powers, & Wallace, 2011) and will not reach full development until sometime during adolescence (Hillock-Dunn & Wallace, 2012). Thus, the unique influence of speech on TWI width is likely fully manifested in individuals who have reached early adulthood. Furthermore, the attenuation of auditory N1/P2 ERP components that is consistently seen in adults in response to audiovisual speech have similarly been found in children who showed attenuated P2 amplitude and an overall growth in this component across development (Knowland et al., 2014). Our integration experience with speech across the lifespan appears to be adaptive which likely results in higher sensitivity to the low-level features present in speech compared to other audiovisual stimuli.

Green and colleagues (2019) further examined this phenomenon through two experiments by pairing simple auditory and visual frequencies (whether within, or outside, human peak sensitivity range) and modulating TWI width using 13 SOAs (0, +/- 50, +/- 100, +/- 150, +/-200, +/-250, or +/-300 ms) and SJ task. To determine if multisensory integration can be modulated depending on whether the perceived bimodal frequencies fall along peak sensitivity range, the study included pure tones composed of frequencies within (400, 800, 1600, and 3200 Hz) and outside (100, 200, 6400, and 12800 Hz) of said range paired with four Gabor patches of frequencies within (1, 3, 6, and 12 cpd) and outside (0.25, 0.50, 25, and 50 cpd) of the peak range.

In Experiment 1a, the visual stimulus (Gabor patch of 3 cpd) was held constant while paired with one of the 8 pure tones at random. For Experiment 1b, the opposite occurred in which a pure tone of 800 Hz was held constant while randomly paired with one of the eight Gabor patches. Results yielded significantly wider TWIs when manipulated stimuli (i.e., sounds for Experiment 1a and Gabor patches for experiment 1b) were within peak sensitivity range. Because the stimuli pairings of Experiments 1a and 1b always included a sound (800 Hz) or Gabor patch (3 cpd) within peak sensitivity range, the researchers revised the paradigm for a second experiment to include frequency pairings that consisted of each modality being paired within and outside peak range. To elaborate, stimuli were paired across modalities (i.e., AV) inside (I) or outside (O) peak sensitivity in the following fashion: auditory inside and visual inside (A_I, V_I), auditory inside and visual outside (A_I, V_O), auditory outside and visual inside (A_O, V_I), and auditory outside and visual outside (A_O, V_O). Subjects were again instructed to perform simultaneity judgement tasks to obtain measures of integration effects. Findings in Experiment 2 indicated that TWI was widest for A_IV_I trials (see: Figure 2.B from Green et al., 2019); additionally, the proportion of simultaneous responses, and RT, was largest when auditory stimuli fell within respective peak sensitivity range, regardless of the paired visual stimuli. Critically, one major implication derived from these results suggests that the different integration patterns due to varying sound and visual spatial frequencies, as well as an interaction between the two modalities, may be occurring very early at primary sensory regions in the brain (Green et al., 2019).

The goal of the current experiment is to follow up on the results found in Green et al. (2019), and to gain insight on the underlying ERP components involved in audiovisual integration of low-level features. We predicted that behavioral responses will be similar to those found by Green and colleagues (2019), in that first, the proportion of simultaneous, or "same," responses should be greatest when frequency pairings are both within their respective peak sensitive range. However, given the dominance of auditory modality when judging temporal information (Repp & Penel, 2002), we also anticipated that the proportion of simultaneous responses may be largest, and reaction times fastest, when auditory modality falls within peak frequency range regardless of frequency in the paired visual modality.

Second, given our natural tendency to integrate visual-leading, AV events more often, we expected to observe a higher proportion of simultaneity judgements for visual-leading trials as a function of SOA, similar to the results showed by Stevenson and Wallace (2013), and thus, the proportion of "same" responses collapsed across frequency parings are anticipated to be largest at the +150ms SOA (visual-first). Although not directly hypothesized, we did not neglect the rule of inverse effectiveness (Holmes, 2007; Meredith & Stein, 1986; Stein and Standford, 2008) and considered the possibility of stimuli residing outside the peak frequency range having the greatest effect on multisensory integration (larger proportion of simultaneous judgements for outside, compared to stimuli inside peak sensitivity range).

At the electrophysiological level, we first expected to observe differences in the sensory ERPs when the stimuli are inside vs. outside the peak sensitivity ranges for each modality (e.g., larger auditory sensory ERPs for A_IV_I and A_IV_O compared to A_OV_I and A_OV_O). If integration influences these early sensory responses, we would also expect to see differential sensory responses when both stimuli are within their respective peak sensitivity ranges compared to when one stimulus is inside and the other is outside (e.g. $A_I V_I$ vs. $A_I V_O$).

Second, we anticipated to observe a distinct difference in ERP waves between trials in which participants perceive the two unisensory events as occurring simultaneously compared to responses that indicate a segmented event. Particularly, it is possible that we may observe speeded and attenuated N1 and P2 components for audiovisual-perceived trials despite our stimuli being non-naturalistic nonspeech stimuli in comparison to those used in the study by Stekelenburg & Vroomen (2007). We expected that this effect, if observed, would be largest when both stimuli are within their respective peak sensitivity ranges ($A_I V_I$) and smallest when both stimuli are outside this range ($A_O V_O$).

Lastly, we expected comparable conclusions to Cecere and colleagues (2017), in that integration effects in the ERP waves may be substantially different between trials in which either the auditory stimuli or the visual stimuli occurred first. We will compare the auditory-first and visual-first integration effects for each of the 300 ms and 150 ms SOAs.

CHAPTER 2

METHODOLOGY

2.1 Participants

A total of 29 healthy participants were recruited at the University of South Carolina; all had normal or corrected-to-normal vision and hearing. Due to excessive noise in some of the data and high artifact rejection rates (< 30% rejected trials), eight subjects were excluded from data analysis, which brought the final number of included participants to 21 (14 females and 7 males). Their ages ranged from 18 to 26 (*M* = 20.76, *SD* = 2.19). Each session was two hours in length and all participants signed informed consent.

2.2 Stimuli and Apparatus

The experiment occurred in a dimly lit, sound-attenuated room. Sound stimuli were presented via speakers located to the left and right of a computer monitor, which presented visual stimuli on a 23-inch display, 65 cm away from the subject. All sounds were projected at 75 dB (fixed to the baseline – 100 Hz). As discussed above, the auditory stimuli were presented as either inside (400, 800, 1600, and 3200 Hz) or outside (100, 200, 6400, and 12800 Hz) the peak sound frequency range. Visual stimuli were represented by Gabor patches (gratings of alternating contrast) either inside (1, 3, 6, or 12 cpd) or outside (0.25, 0.50, 25, or 50 cpd) peak spatial frequency ranges (see Appendix A for spectrum of frequencies). Conditions were represented in stimulus frequency pair identical to Experiment 2 by Green et al., 2019 (i.e., A_I and V_I, A_I and V_O, A_O and V_I, and A_O and V_O). To maximize the number of trials for ERP analysis, the number of SOAs were reduced to 5: 0, +/- 150 ms, and +/- 300 ms. These SOAs were chosen because their timing produces reliable integration (0/simultaneous), reliable segregation (+/- 300 ms), and are near the temporal limit of integration (+/- 150 ms) and thus produce approximately equal numbers of trials where integration does and does not occur. Performance as a function of SOAs, as well as ERPs evoked in each stimulus-pair condition, were examined to explore how stimulus-induced changes in TWI width relate to changes in neural activity.

2.3 Procedure

Procedures will be identical to Experiment 2 presented in Green et al. (2019), with the exception of fewer SOAs and the recording of EEG. After receiving general task instructions, participants practiced until they felt comfortable starting the actual experiment (typically between 10-15 trials). Furthermore, subjects were instructed to fixate their gaze on a fixation point centered on the screen and encouraged to blink only when necessary, in between trials. Participants responded manually by clicking either the left (“same”) or right (“different”) buttons on a mouse pad during the SJ task to determine whether the auditory and visual stimuli were perceived to occur at the same, or at different times. In total, we presented 20 blocks of 60 trials (1200 trials), with self-paced break periods in between blocks. Each trial was self-paced, meaning that next trial did not appear until participants’ response was recorded. Each response was followed by an inter-trial interval ranging from 1000-1500 ms prior to the next trial (see Figure 1).

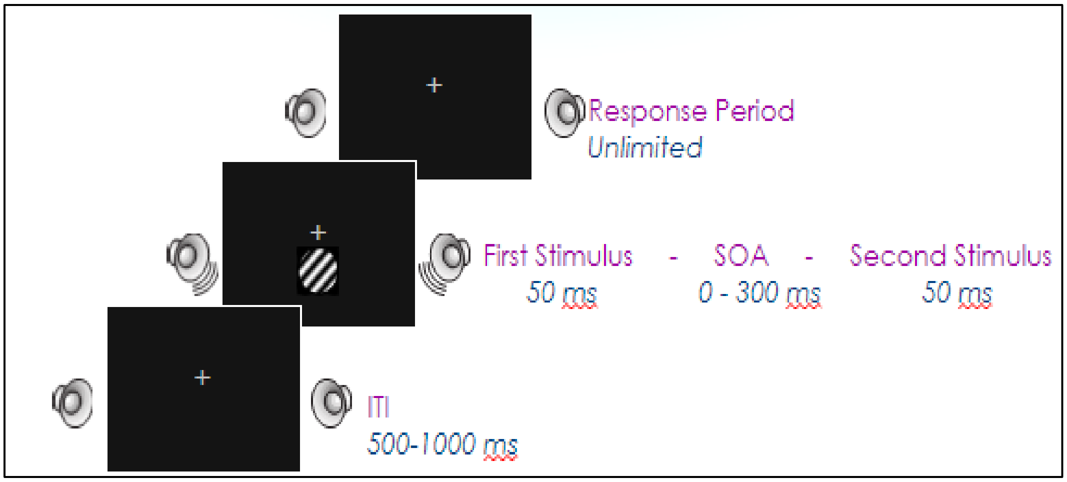


Figure 2.1. Visual example of a trial that participants encountered. Inter-trial interval (ITI) period randomly varied between 500 – 1000ms and was followed by the audiovisual stimuli pairings separated by SOA.

2.4 EEG Recording and Analysis

EEG was recorded with a 500 Hz sampling rate from 32 locations (AF3, AFz, AF4, FC5, FC3, FC1, FCz, FC2, FC4, FC6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, PO11, PO9, PO7, PO3, POz, PO4, PO8, PO10, PO12, I1, Iz, I2, IIz, and M1) using the international 10-20 system for electrode placement. Left mastoid (M1) was used as the reference electrode during recording and data was digitally re-referenced to the average activity of the left and right mastoids during analysis. Two horizontal electrooculogram (HEOG) electrodes were used to measure and record horizontal eye movement to aid in eye movement artifact removal during analysis. All trials containing blinks or eye movement artifacts were removed from further analysis, and participants who had more than 30% rejected trials were not used in subsequent analyses. We used EEGLAB and ERPLab toolbox extensions of the programming software, MATLAB, to analyze the raw EEG data. Infinite impulse response (IIR) Butterworth filter was used with high- and low-pass filtering at .01 and 30 Hz, respectively.

Averages were created for each condition for each participant (A_IV_I, A_IV_O, A_OV_I, and A_OV_O for each SOA). For the 150 ms and 300 ms SOA, separate averages were created for auditory-first and visual first trials. All ERPs were baseline corrected using the 200 ms prior to the first stimulus in the pair. For each SOA we compared sensory evoked potentials; visual responses at occipital electrodes (PO7 and PO8) and auditory responses at fronto-central electrodes (FCz) across inside/outside pairings to determine whether sound or spatial frequency influences responses. In addition, separate averages were created for each condition based on the participant's response (i.e., same or different). This allowed us to compare neural responses to physically identical stimuli when they were integrated (same response) compared to when they were not (different response). This will be particularly relevant for the +/- 150ms SOAs, which are near the limit of the TWI and should produce approximately equal same/different responses for most participants. Mean amplitudes and peak latencies of the ERPs waveforms were compared using within-subjects ANOVAs.

CHAPTER 3

RESULTS

3.1 Behavioral Results

General linear models were used to address the aforementioned behavioral hypotheses. Bonferroni corrections were applied to all post-hoc analyses. First, a 2 x 2 x 5 (auditory modality vs. visual modality vs. SOA) repeated measures ANOVA with two (inside or outside) and five (0ms, +/- 150ms, and +/- 300ms) levels of modalities and SOA, respectively, revealed significant outcomes. A main effect of auditory modality [$F(20, 1) = 7.874$, $p = .011$, $\eta^2 = 0.002$] and SOA [$F(20, 4) = 34.936$, $p < .001$, $\eta^2 = 0.602$] emerged. A main effect of visual modality ($p = .992$), or any other interactions (all p's > .159) were not found (see Table B.1 for complete ANOVA results). Post-hoc analysis revealed only partial support for our behavioral hypotheses. A greater proportion of simultaneous responses for A_IV_I compared to A_OV_O trials were not found for any of the 5 SOA conditions (all p's > .143). This outcome also negates any potential instance of inverse effectiveness. However, more instances of simultaneity were reported for trials in which auditory stimuli were inside, compared to outside, of the peak sensitivity range [$t(20) = 2.806$, $p = .011$)]. This finding occurred regardless of paired visual stimuli, indicating auditory dominance as previously predicted in behavioral hypothesis two. The same analysis was conducted on RT and a main effect of SOA [$F(20, 4) = 41.612$, $p < .001$, $\eta^2 = 0.519$] emerged. A post-hoc analysis showed that participants were fastest to respond during the 0ms SOA compared to all other SOAs [0ms vs. -150ms: $t(20) = -5.094$, $p < .001$; 0ms vs. +150ms: $t(20) = -6.337$, $p < .001$; 0ms vs. -300ms: $t(20) = -10.986$, $p < .001$; 0ms vs. +300ms: $t(20) = -10.855$, $p < .001$]. Additionally, an interaction between auditory modality and SOA [$F(20, 1) = 3.568$, $p = .010$, $\eta^2 = 0.014$] was found. Post-hoc analysis revealed that reaction times for auditory-inside stimuli at the 0ms SOA, regardless of visual frequency pairing, at were significantly faster than all other SOAs in which auditory modality was inside or outside (all *p*'s < .001), except for 0ms auditory-outside trials ($p = 1.000$). A main effect of auditory ($p = .530$) or visual modality ($p = .071$), or any other interactions (all *p*'s > .328), were not found. Although no differences between auditory-inside compared to auditory-outside, regardless of visual stimuli pairings, were found [$t(20) = -0.639$, $p = .530$], the results on RT suggest an effect of auditory dominance that is likely modulated by SOA (see Appendix C).

Second, a one factor (SOA) repeated measures ANOVA revealed significant findings, $F(20, 4) = 34.897$, $p < .001$, $\eta^2 = 636$. Instances of simultaneity as reflected by the proportion of simultaneous responses made, regardless of frequency condition, were greater for the 0ms SOA in comparison to the +150ms ($p = .002$), and +/- 300ms (all *p*'s < .001) SOAs, but not the -150ms SOA ($p = 1.000$). The -150ms SOA, regardless of frequency condition, also resulted in larger instances of integration in comparison to the +150ms ($p = .001$) and + 300ms ($p < .001$) but not -300ms ($p = 1.000$) SOAs (see Figure 2).

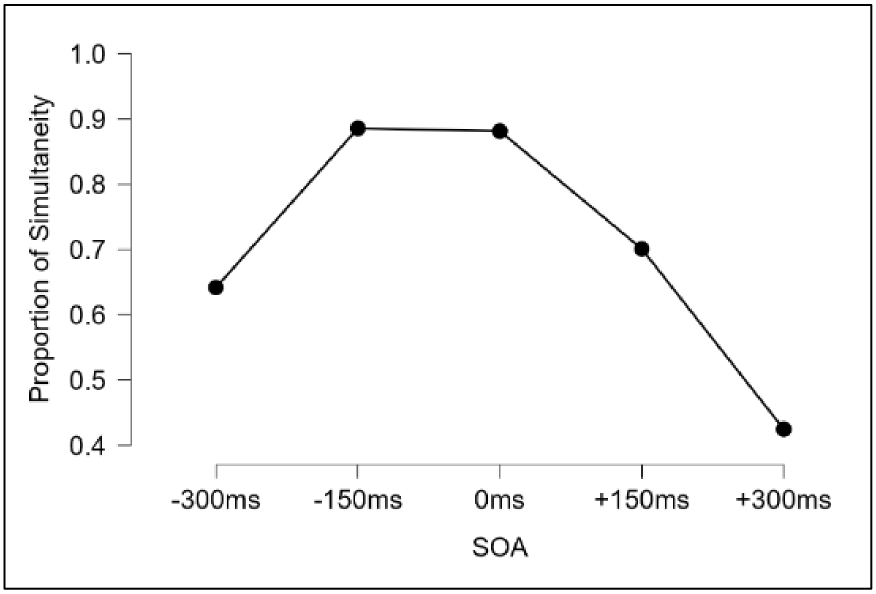


Figure 3.1. Proportion of simultaneous responses at each SOA collapsed across frequency condition.

3.2 Electrophysiological Results

For sensory ERP analyses, mean amplitude and peak latency comparisons of "same" judgements were executed; because we are the first, to our knowledge, to investigate ERPs in this context, the analyses were primarily exploratory. However, due to the findings presented by Stekelenburg and Vroomen (2007), we aimed our investigation towards the N1 and P2 ERP components. To examine potential differences in the sensory ERPs for each modality and to determine whether ERPs for inside compared to outside frequencies are larger, a 2 x 2 (auditory modality vs. visual modality) repeated measures ANOVA with two levels (inside or outside) was conducted for each SOA at sensory electrode sites (FCz for auditory ERPs and PO8 for visual ERPs). Bonferroni corrections were applied to all post-hoc analyses.

3.2.1 0ms SOA

For the 0ms SOA auditory N1 component, a main effect of visual modality [$F(20, 1) = 7.852$, $p = .011$, $\eta^2 = 0.113$], but not auditory modality [$F(20, 1) =$, $p = .488$] was found. Post-hoc analysis revealed significant attenuation for auditory N1 component amplitude for A_IV_I compared to A_IV_O trials [$t(20) = 3.337$, $p = .011$; see Figure 3] but not in comparison to any other condition (all p’s > .104). Auditory N1 latency comparisons did not reveal any main effects or interactions (all p’s > .064). Although we did not find any main effects or interactions of modality on auditory P2 amplitude (all p’s > .117), a main effect of visual modality [$F(20, 1) = 5.541$, $p = .029$, $\eta^2 = 0.084$], but not auditory modality [$F(20, 1) = 0.355$, $p = .558$], was found for auditory P2 latency. However, post-hoc analysis did not reveal any significant comparisons (all p’s > .241). At PO8 electrode site, visual N1 amplitude did not reveal any significant outcomes, only a main effect of visual modality approached significance ($p = .053$; all other p’s > .103). However, visual N1 latency showed a main effect of visual modality [$F(20, 1) = 5.665$, $p = .027$, $\eta^2 = 0.152$] but no significant post-hoc comparisons (all p’s > .113). Additionally, main effects of visual modality on visual P2 amplitude [$F(20, 1) = 14.131$, $p = .001$, $\eta^2 = 0.334$; all other p’s >.227] and latency [$F(20, 1) = 8.747$, $p = .008$, $\eta^2 = 0.241$; all other p’s >.226] were found. Regarding visual P2 amplitude, several significant post-hoc comparisons were found. First, P2 amplitude for A_IV_I trials was significantly enhanced in comparison to A_IV_O [$t(20) = 2.993$, $p = .035$] and A_OV_O [$t(20) = 3.254$, $p = .018$] trials. Similarly, P2 amplitude for A_OV_I trials were enhanced in comparison to A_IV_O [$t(20) = 3.557$, $p = .008$] and A_OV_O [$t(20) = 3.950$, $p = .003$] trials. Post-hoc comparisons for visual P2 latency revealed only an increase in peak latency for A_OV_I compared to A_OV_O trials [$t(20) = 3.209$, $p = .021$]. At the PO7 electrode site, similar findings emerged; a main effect visual modality on N1 latency [$F(20, 1) = 5.233$ $p = .033$, $\eta^2 = 0.113$; all other p's >.332], P2 amplitude [$F(20, 1) = 9.243$, $p = .006$, $\eta^2 = 0.203$; all other p's >.466], and P2 latency [$F(20, 1) = 13.368$ $p = .002$, $\eta^2 = 0.113$; all other p's >.117] were found. Post-hoc analyses showed enhanced P2 amplitude for A_IV_I compared to A_OV_O [$t(20) = 3.005$, $p = .031$] and A_OV_I compared to A_OV_O [$t(20) = 2.903$, $p = .038$] trials; additionally, P2 latency occurred earlier for A_IV_I compared to A_OV_O trials [$t(20) = -4.001$, $p = .002$]. See Figure 4.

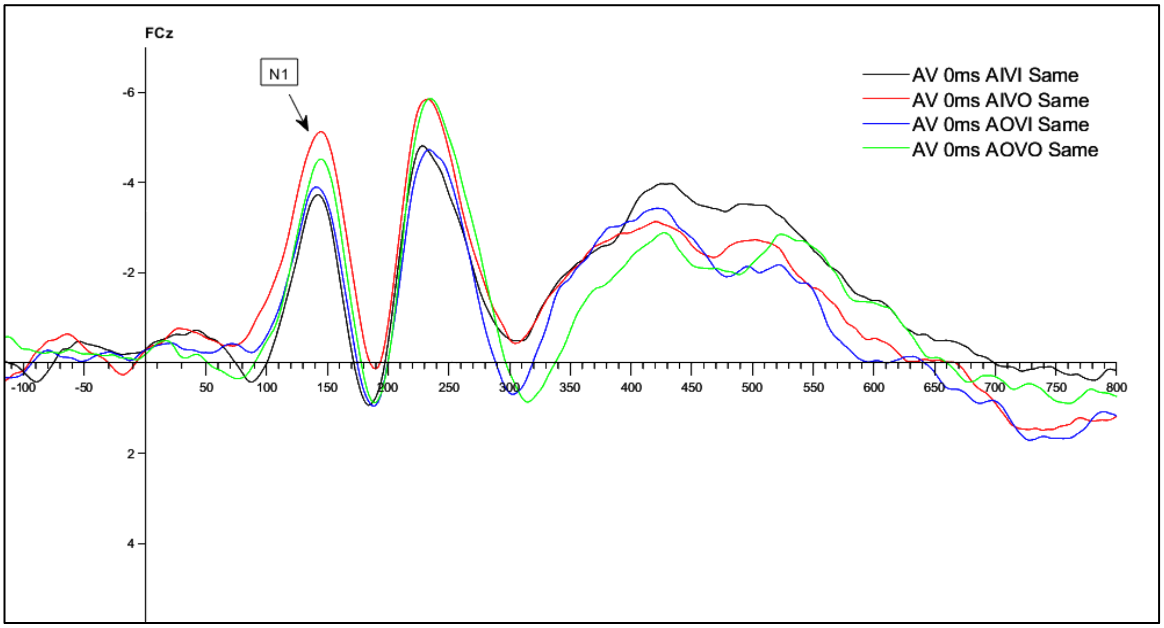


Figure 3.2. Plot of auditory ERP wave at electrode site FCz when both visual and auditory stimuli occurred simultaneously (0ms SOA). The A_IV_I N1 (black) is shown to be attenuated in comparison to the A_IV_O N1 (red).

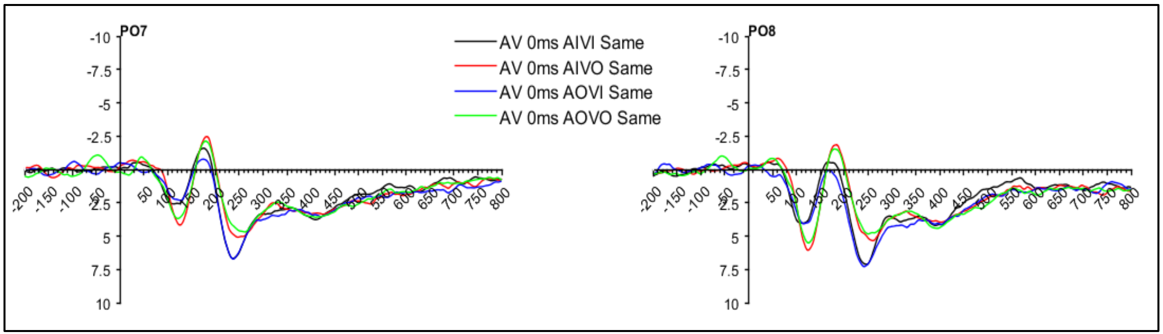


Figure 3.3 Plot of visual ERP waves at electrode sites PO7 and PO8 for simultaneously occurring visual and auditory stimuli (0ms SOA). The A_IV_I (black) and A_OV_I (blue) P2 ERPs are shown to be enhanced.

3.2.2 -150ms SOA

At the -150ms SOA, only the auditory P2 component was found to be significant. A main effect of auditory modality on amplitude [$F(20, 1) = 6.073$, $p = .023$, $\eta^2 = 0.098$; all other p's >.370] was found. Post-hoc analysis did not reveal any significant comparisons (all p's > .133). As for P2 latency, a main effect of auditory modality was found [$F(20, 1) = 10.820$, $p = .004$, $\eta^2 = 0.102$; all other p's >.091] with post hoc comparisons showing speeded latency for A_IV_O compared to A_OV_I trials [$t(20) = -3.362$, $p = .011$]. For visual components at PO8 sites, a main effect of visual modality on N1 latency was found [$F(20, 1) = 14.986$, $p < .001$, $\eta^2 = 0.228$; all other p's >.683], with several post-hoc findings. First, N1 latency was significantly speeded for A_IV_I compared to A_IV_O [$t(20) = 2.932$, $p = .033$] and A_OV_O [$t(20) = 3.205$, $p = .016$] trials. Similarly, N1 latency was also speeded for A_OV_I trials but only in comparison to A_OV_O trials [$t(20) = 2.843$, $p = .042$]. Additionally, we found a main effect of visual modality on P2 amplitude [$F(20, 1) = 24.299$, $p < .001$, $\eta^2 = 0.410$; all other p's > .293]. Further, a post-hoc analysis revealed significantly larger P2 amplitude during trials in which the visual stimuli were inside peak frequency range, regardless of the paired auditory stimuli [A_IV_I vs. A_IV_O: $t(20) = 3.710$, $p = .005$; A_IV_I vs. A_OV_O: $t(20) = 4.698$, $p < .001$; A_OV_I vs. A_IV_O: $t(20) = 3.732$, $p = .004$; A_OV_I vs. A_OV_O: $t(20) = 4.801$, $p < .001$; see Figure 5]. No main effects or interactions were found for visual P2 latency (all p's > .452). At PO7 sites, we found similar results, in that, a main effect of visual modality on N1 latency [$F(20, 1) = 7.944$, $p = .011$, $\eta^2 = 0.128$; all other p's > .166], P2 amplitude [$F(20, 1) = 15.990$, $p < .001$, $\eta^2 = 0.268$; all other p's > .156], and P2 latency [$F(20, 1) = 20.161$, $p < .001$, $\eta^2 = 0.192$; all other p's > .730] emerged. Post-hoc comparisons for N1 latency and P2 amplitude revealed earlier latency for A_OV_I compared to A_IV_O [$t(20) = -3.163$, $p = .021$] trials and enhanced amplitude for A_OV_I compared to A_IV_O [$t(20) = 3.729$, $p = .004$] and A_OV_O [$t(20) = 4.161$, $p = .001$] trials, respectively. P2 latency occurred earlier for A_IV_I compared to A_OV_O [$t(20) = -3.080$, $p = .023$] and A_OV_I compared to A_OV_O [$t(20) = -2.979$, $p = .030$] events; see Figure 5.

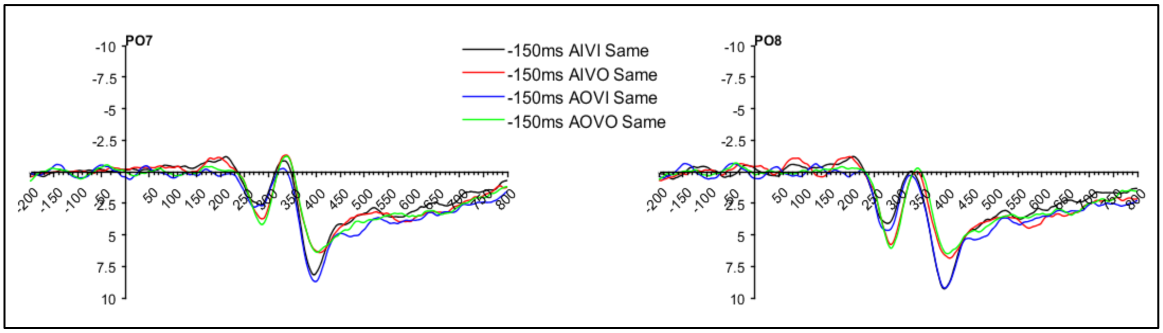


Figure 3.4. Plot of visual ERP waves at electrode sites PO7 and PO8 for auditory-first occurring audiovisual stimuli separated in time by 150ms (-150ms SOA). The A_IV_I (black) and A_OV_I (blue) P2 ERPs are shown to be enhanced.

3.2.3 +150ms SOA

The +150ms SOA revealed a main effect of visual modality on auditory N1 latency [*F*(20, 1) = 7.343, *p* = .013, η^2 = 0.130; all other *p*’s >.486]; however, post-hoc comparisons did not reveal anything significant (all *p*’s > .084). No main effects or interactions were found for auditory N1/P2 amplitude or P2 latency (all *p*’s >.083). A main effect of visual modality on visual N1 latency, but not amplitude (all *p*’s > .083), was marginally significant [*F*(20, 1) = 4.360, *p* = .050; all other *p*’s > .443] at PO8 sites with no significant post-hoc comparisons (all *p*’s > .230). Additionally, visual P2 amplitude resulted in a main effect of visual modality [*F*(20, 1) = 15.590, *p* < .001, η^2 = 0.278; all other *p*’s >.605], with post-hoc analysis showing enhanced amplitude for trials containing visual stimuli within peak frequency range regardless of paired auditory stimuli [A_IV_I vs. A_IV_O: *t*(20) = 2.880, *p* = .040; A_IV_I vs. A_OV_O: *t*(20) = 3.093, *p* = .023; A_OV_I vs. A_IV_O: *t*(20) = 3.396, *p* = .010; A_OV_I vs. A_OV_O: *t*(20) = 3.500, *p* < .007; see Figure 6]. No main effects or interactions were found for visual P2 latency (all *p*’s > .143) at PO8 sites. Furthermore, we found a main effect of visual modality on N1 latency [*F*(20, 1) = 13.003, *p* = .002, η^2 = 0.178; all other *p*’s >.353] at PO7 sites with post-hoc analysis showing speeded latency for A_IV_I compared to A_OV_O events [*t*(20) = -3.462, *p* = .008]. Additionally, a main effect of visual modality on P2 amplitude [*F*(20, 1) = 6.139, *p* = .022, η^2 = 0.118; all other *p*’s >.607] and latency [*F*(20, 1) = 7.316, *p* = .014, η^2 = 0.278; all other *p*’s >.217], with post-hoc comparisons revealing only speeded latency for A_IV_I compared to A_IV_O [*t*(20) = -2.864, *p* = .040] trials; see Figure 6.

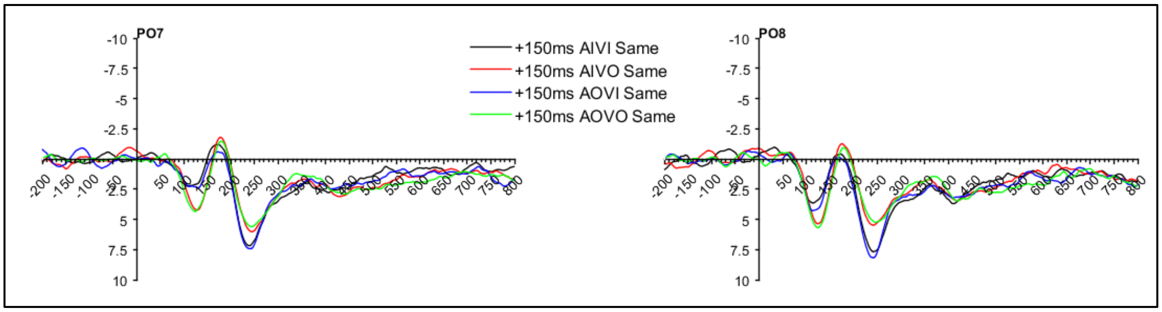


Figure 3.5. Plot of visual ERP waves at electrode sites PO7 and PO8 for visual-first occurring audiovisual stimuli separated in time by 150ms (+150ms SOA). The A_IV_I (black) and A_OV_I (blue) P2 ERPs are shown to be enhanced.

3.2.4 -300ms SOA

At the -300ms SOA, a main effect of auditory modality on auditory P2 amplitude emerged [$F(20, 1) = 10.657$, $p = .004$, $\eta^2 = 0.085$; all other p's >.180]. Post-hoc comparisons showed significant enhancement of P2 amplitude for A_OV_I when compared to A_IV_O trials [$t(20) = 2.992$, $p = .030$]. Additionally, a main effect of visual modality was found on auditory P2 latency [$F(20, 1) = 5.055$, $p = .036$, $\eta^2 = 0.065$; all other p's >.147], with no significant post-hoc analysis outcome (all p's > .073). There were no significant outcomes for auditory N1 components, only a main effect of auditory modality on N1 amplitude approaching significance ($p = .068$; all other p's > 150). Moreover, several findings emerged for visual components at PO8/PO7 sites. First, at PO8, a main effect of auditory modality on N1 amplitude emerged [$F(20, 1) = 8.540$, $p = .008$, $\eta^2 = 0.058$; all other p's >.682] with no significant post-hoc comparisons (all p's > .381). Second, a main effect of visual modality on N1 latency was found [$F(20, 1) = 9.453$, $p = .006$, $\eta^2 = 0.122$; all other p's >.565], but no significant post-hoc comparisons emerged (all p's > .092). Lastly, a main effect of auditory [$F(20, 1) = 10.234$, $p = .005$, η^2

= 0.045] and visual modality [*F*(20, 1) = 8.025, *p* = .010, η^2 = 0.208], but no interaction (*p* = .930), was found for P2 amplitude. Post-hoc comparisons only showed significantly enhanced P2 amplitude for A_OV_I compared to A_IV_O trials [*t*(20) = 3.853, *p* = .004]. No main effects or interactions (all *p*'s > .508) were found for P2 latency. At PO7 site, a main effect of auditory modality on N1 amplitude was found [*F*(20, 1) = 6.611, *p* = .018, η^2 = 0.065, all other *p*'s = 423], with visual modality effects on N1 latency approaching significance (*p* = .054). A main effect of auditory [*F*(20, 1) = 5.344, *p* = .032, η^2 = 0.055] and visual [*F*(20, 1) = 11.141, *p* = .003, η^2 = 0.184] modality, but no interaction (*p* = .426) were found on P2 amplitude. Post-hoc comparisons showed enhanced amplitude for A_IV_I [A_IV_I vs. A_IV_O: *t*(20) = 3.105, *p* = .021] and A_OV_I [A_OV_I vs. A_IV_O: *t*(20) = 4.052, *p* = .001] events. P2 latency showed a main effect of visual modality only [*F*(20, 1) = 7.557, *p* = .012, η^2 = 0.083; all other *p*'s > .483]. See Figure 7.

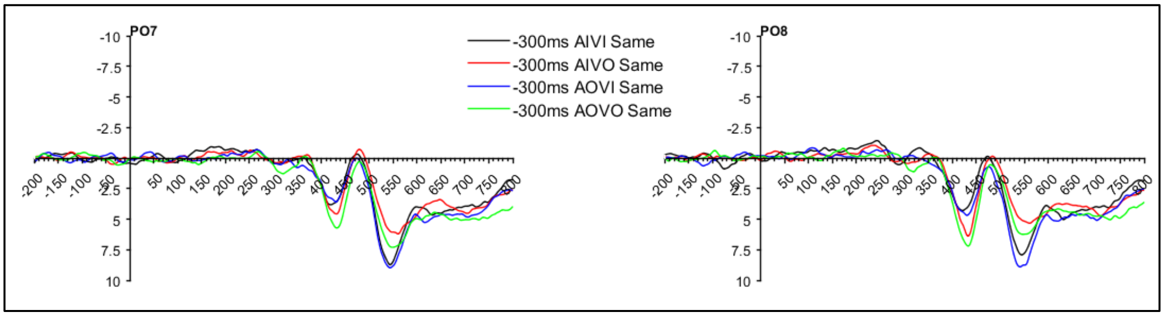


Figure 3.6. Plot of visual ERP waves at electrode sites PO7 and PO8 for auditory-first occurring audiovisual stimuli separated in time by 300ms (-300ms SOA). The A_IV_I (black) and A_OV_I (blue) P2 ERPs are shown to be enhanced.

3.2.5 +300ms SOA and Inverse Effectiveness

At the +300ms SOA, several major findings were observed. A main effect of visual modality on visual N1 latency was found [$F(20, 1) = 24.789$, $p < .001$, $\eta^2 = 0.258$; all other *p*'s > .404]. Post-hoc analysis revealed speeded latency for trials containing visual stimuli within peak frequency range regardless of paired auditory stimuli [A_IV_I vs. A_IV_O: $t(20) = -2.880$, $p = .039$; A_IV_I vs. A_OV_O: $t(20) = -3.073$, $p = .023$; A_OV_I vs. A_IV_O: $t(20) = -4.231$, $p < .001$; A_OV_I vs. A_OV_O: $t(20) = -3.215$, $p = .016$]. For visual ERPs at PO8, we found a main effect of visual modality on visual P2 amplitude [$F(20, 1) = 17.965$, $p < .001$, $\eta^2 = 0.289$; all other *p*'s > .399], with post-hoc comparisons showing enhanced P2 amplitude for visual-inside trials regardless of paired auditory frequencies [A_IV_I vs. A_IV_O: $t(20) = 3.930$, $p = .002$; A_IV_I vs. A_OV_O: $t(20) = 3.443$, $p = .008$; A_OV_I vs. A_IV_O: $t(20) = 3.141$, $p = .019$; A_OV_I vs. A_OV_O: $t(20) = 2.912$, $p = .036$]. No significant findings emerged among visual N1 amplitude, P2 latency, or auditory N1/P2 components (all *p*'s > .081). At PO7, a main effect of visual modality on N1 latency [$F(20, 1) = 6.080$, $p = .023$, $\eta^2 = 0.109$; all other *p*'s > .193], P2 amplitude [$F(20, 1) = 37.461$, $p < .001$, $\eta^2 = 0.224$; all other *p*'s > .939], and P2 latency [$F(20, 1) = 7.396$, $p = .013$, $\eta^2 = 0.102$; all other *p*'s > .302] were found. Post-hoc comparisons were only significant for P2 amplitude, showing enhancements for A_IV_I and A_OV_I trials [A_IV_I vs. A_IV_O: $t(20) = 4.075$, $p = .001$; A_OV_I vs. A_OV_O: $t(20) = 3.959$, $p = .002$]; see Figure 8. Moreover, we did not find any significant evidence of inverse effectiveness, as A_OV_O ERPs were not largest at any sensory electrode site across SOAs.

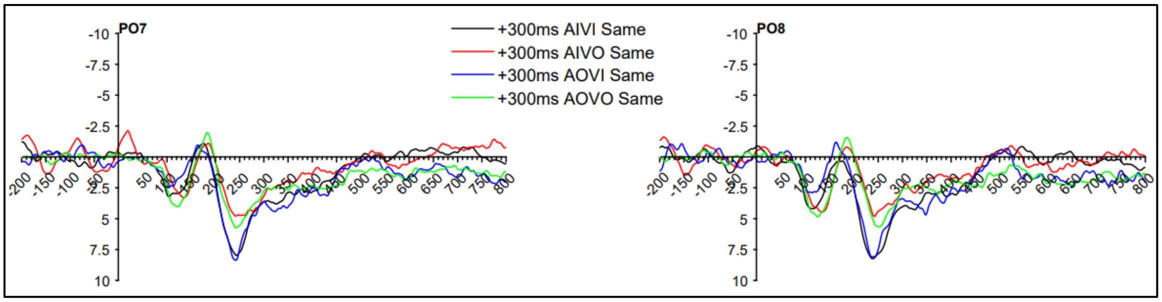


Figure 3.7. Plot of visual ERP waves at electrode sites PO7 and PO8 for visual-first occurring audiovisual stimuli separated in time by 300ms (+300ms SOA). The A_IV_I (black) and A_OV_I (blue) P2 ERPs are shown to be enhanced.

3.2.6 Integration vs. Non-integration

To investigate differences in the ERPs between trials in which participants integrated information compared to when they did not (i.e., “same” vs. “different” judgments, or same/different) we first compared ERPs across frequency at each SOA and followed up with an A_IV_I / A_OV_O comparison. Paired samples t-tests were used to compare amplitude and latency measures. When comparing ERPs of same/different across frequency at the 0ms SOA, we did not find any differences in auditory or visual N1/P2 amplitude or latency for integrated compared to non-integrated trials (all *p*’s > .235). However, at the -150ms SOA, we found a significant difference of auditory P2 latency between integrated and non-integrated trials [$t(20) = 2.252$, $p = .036$], with integrated trials showing larger latency. No differences were found for auditory N1/P2 amplitude or N1 latency (all *p*’s > .394). For visual components, only the N1 amplitude resulted in a marginally significant difference between integrated and non-integrated trials [$t(20) = 2.085$, $p = .050$], with integrated trials having attenuated N1 amplitude. No significant findings emerged for visual N1/P2 latency; only P2 amplitude reached near significance ($p = .066$; all other *p*’s > .192). As for the +150ms SOA, across all auditory and visual N1/P2 components, only auditory N1 amplitude reached near significance (*p* = .059; all other *p*'s >.090). For temporal differences of 300ms, two outcomes were observed. First, at the -300ms SOA, N1 amplitude difference between integrated and non-integrated was found trials [*t*(20) = -2.214, *p* = .039], showing attenuated N1 amplitude for integrated trials. Second, visual N1 latency difference was significant at the +300ms SOA, showing speeded latency for integrated compared to non-integrated trials [*t*(20) = 2.413, *p* = .026]. trials [*t*(20) = 2.252, *p* = .036].

To examine whether a larger difference in amplitude and latency between same/different occurred as a function of frequency, we compared "same" vs. "different" responses at A_IV_I, A_IV_O, A_OV_I, and A_OV_O trials. Importantly, we analyzed at each frequency condition only at the +150ms and -300ms SOAs as these included relatively equal "same" and "different" judgements. At the +150ms SOA, we only found speeded auditory N1 latency during A_IV_O trials [*t*(20) = 2.236, *p* = .037], as well as enhanced auditory N1 amplitude during A_OV_I trials [*t*(20) = -3.463, *p* = .002]. However, there were no significant differences of amplitude or latency between same/different responses at either A_IV_I or A_OV_O trials (all *p*'s > 108). However, at the -300ms SOA, several findings emerged for visual components in A_IV_I trials. At PO8 electrode site, N1 amplitude was significantly enhanced for integrated compared to non-integrated trials [*t*(20) = -2.972, *p* = .008]. Additionally, P2 amplitude [*t*(20) = -2.248, *p* = .036] and latency [*t*(20) = 2.363, *p* = .028] were shown to be attenuated for integrated trials (see Figure 9). At PO7 electrode site, similar findings emerged in enhanced N1[*t*(20) = -3.096, *p* = .006] and suppressed P2 [*t*(20) = -2.616, *p* = .017] amplitude, but not P2 latency [*p* = .102], for integrated trials. Additionally, for A_IV_O trials, we found speeded latency for visual P2 latency [$t(20) = 2.763$, $p = .012$]. No significant findings emerged for A_IV_I auditory N1/P2 components, nor for A_OV_I /A_OV_O auditory or visual N1/P2 components (all p's > .079).

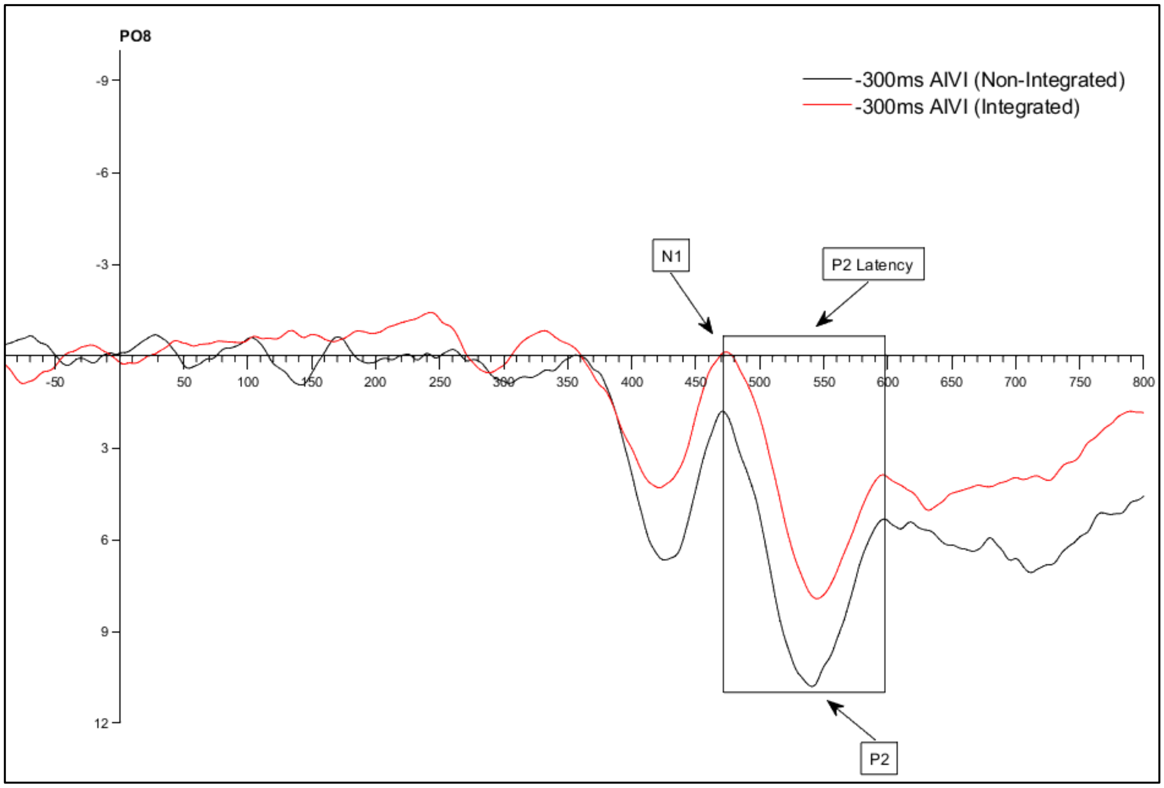


Figure 3.8. Plot of visual ERP wave at electrode site PO8 for auditory-first occurring events (-300ms SOA) between simultaneous (red) and non-simultaneous (black) responses. N1 amplitude was enhanced and proceeded by speeded P2 latency and enhanced amplitude.

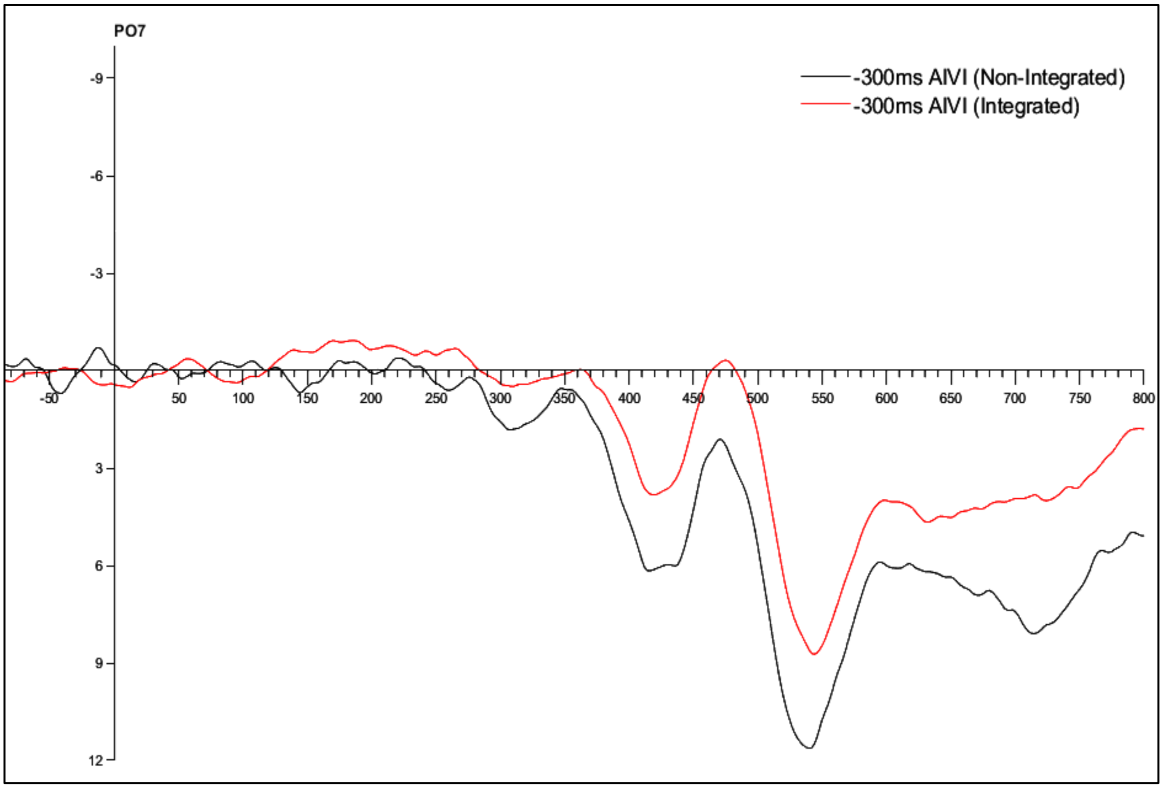


Figure 3.9. Plot of visual ERP wave at electrode site PO7 for auditory-first occurring events (-300ms SOA) between simultaneous (red) and non-simultaneous (black) responses. N1 amplitude was enhanced and proceeded by enhanced P2 amplitude.

3.2.7 A-First vs. V-First

Third, to examine differences in ERP waves between auditory-first and visual-first trials, we computed the difference between mean amplitude of modality-specific N1 and P2 at +/-150 and +/- 300ms SOAs. More specifically, to account for differences in ERP timing across SOAs, we computed the difference between the mean amplitudes of the auditory N1 and P2 and between the visual N1 and P2 for each SOA and condition. We then compared the modality specific responses for auditory-first and visual-first trials. Using a paired-samples t-test, we found only one significant outcome which occurred at the -300ms SOA, in that, auditory N1/P2 amplitude differences varied between auditory- and visual-first trials [$t(20) = 3.934$, $p < .001$; all other *p*’s > .208; see Appendix D].

CHAPTER 4

DISCUSSION

Crossmodal integration involves combining multiple unisensory streams of information to form a single perceptual event. While factors such as spatial proximity (Chen, 2013; Spence, 2013; Wallace et al., 1998) and stimulus effectiveness (i.e., Inverse Rule; Holmes, 2007; Meredith & Stein, 1986; Stein & Standford 2008) are highly critical in determining the outcome of integration, here we focused on the temporal aspects of crossmodal binding. In particular, we aimed to gain a better understanding of the neural properties involved in the underlying, low-level components of speech, which has continuously shown to evoke the largest effects of multisensory integration (Spence & Squire, 2003; Stevenson & Wallace, 2013; Vatakis & Spence, 2010).

4.1 Behavioral Implications

At the behavioral level, our results were partially consistent with those of Experiment 2 by Green and colleagues (2019). We found a greater proportion of simultaneous responses for events in which the auditory stimuli were within peak sensitivity range, independent of the frequency in paired visual stimuli. Not only were we able to replicate this finding of Green et al. (2019), we similarly demonstrated the natural dominance of the auditory system during tasks that prioritize temporal information (Repp & Penel, 2002). Since auditory stimuli residing inside the peak sensitivity range elicited the strongest multisensory effects, it is likely that this outcome can serve as a partial explanation for why speech stimuli is a “special” crossmodal binding process. Independent of frequency, it was evident that the largest integration effects occurred when the bimodal stimuli happened either simultaneously (i.e., 0ms SOA) or when the auditory preceded the visual stimuli by 150ms (i.e., -150ms SOA), which was consistent with the reports by Green et al. (2019). Similarly, participants’ response speed was fastest for simultaneous trials compared to all other SOA conditions, suggesting perhaps that sometimes additional evaluative processes occurred when stimuli were presented some time apart, even when the auditory preceded the visual stimuli by 150ms. It is important to note that we did not find larger integration effects for visual-first occurring trials, as reported by Stevenson & Wallace (2013). Although some of our stimuli contained low-level features of speech, beep sounds and Gabor patches are not typically occurring stimuli in the environment and thus, the visual information presented here did not elicit predictive cues for, or any semantic relationship with, the proceeding auditory stimuli, as occurs in speech (Calvert & Thesen, 2004; Chen and Spence, 2017; Spence & Squire, 2003).

4.2 Electrophysiological Implications

At the electrophysiological level, it has been demonstrated that auditory N1 and P2 ERP components experience suppression in amplitude and latency when sound is accompanied by visual information during speech (van Wassenhove et al., 2005) and natural, non-speech events (Stekelenburg & Vroomen, 2007). Our findings partially support those of Stekelenburg & Vroomen (2007) and van Wassenhove et al. (2005), in that we observed attenuated auditory-evoked N1, but not P2, for A_IV_I events during physically synchronous events (i.e., 0ms SOA). However, because this attenuation occurred only in comparison to A_IV_O events, we cannot conclude that audiovisual events composed of highly sensitive frequencies results in attenuation when compared to frequencies outside of said range. Rather, it seems that this process is modulated by the paired visual stimuli and attenuation occurs only when the paired stimuli are not within peak range. It is possible that some of the anticipated effects (i.e., attenuation of N1/P2 amplitude and latency) may not have emerged simply due to our stimuli. Given the lack of semantic relationship between the tone beeps and Gabor patches, it is likely that there is no direct multisensory benefit that one can offer the other, nor any predictive cueing, as occurs in speech. Additionally, humans become accustomed to audiovisual speech as verbal interactions typically occur on a daily basis throughout the lifespan, making it a natural, consistently occurring event. Although some of the stimuli used in this experiment were low-level features that make up audiovisual speech information, tone beeps and Gabor patches are not natural events that humans are accustomed to (van Wassenhove et al., 2005). Therefore, this may serve as a possible explanation as to why our results were only partially aligned with those found by Stekelenburg & Vroomen (2007) and van Wassenhove and colleagues (2005).

Although we were not able to replicate auditory N1/P2 component attenuation, we found strikingly opposite effects at visual sensory components consistently across all SOA conditions. Whether stimuli were displayed simultaneously (0ms SOA), or some time apart (+/- 150/300ms SOAs), visual N1 latency was consistently sped up and P2 amplitude was remarkably enhanced for audiovisual pairings that included visual stimuli residing within peak sensitivity range (V_I). Further, this occurred bi-laterally, as we found nearly identical sensory ERPs at PO7 and PO8 electrode sites which indicates that this finding is not hemisphere-specific. The main effect of visual frequency appears to be a function of the sensitivity to those frequencies that are present in speech and is possibly modulated by the paired auditory frequency. However, the effect seemed slightly weakened for the +/-300ms SOAs which could have two possible explanations. The TWI is narrower for simple stimuli in comparison to more complex speech (van Wassenhove et al., 2005) and ecologically valid (Stekelenburg & Vroomen, 2007) stimuli, and thus, integration effects could be less prominent. However, a second possibility may reside in our sample size lacking in adequate quantity and therefore, contributing to insignificant differences. Nonetheless, these findings demonstrate potential insight to the specialness of speech in which larger sensory ERPs were elicited for those visual frequencies that are prominent in the faces of speakers.

Overall, perhaps enhanced P2 amplitude for simple, nonspeech stimuli may reflect later stage processing that is unique to these specific types of frequencies when they onset some time apart. Given that +/-150/300ms SOAs revealed similar findings, it is possible that there may not be a difference in neural activity between them. In fact, our findings in large did not support those of Cecere et al. (2017), in that, our comparisons between N1/P2 amplitude differences of the +/-150ms SOAs did not differ. It could be argued that neural activity of integration differs between AV- and VA-300ms SOAs at auditory sensory sites, as represented by our observed comparison between N1/P2 amplitude differences for AV and VA events. However, it is unclear whether this difference is perhaps due to early activity onset from primary visual cortex (see Appendix D) or because of some noise due to our low sample size. One possible consideration for inconsistency between our findings and those of previous work could be differences in SOAs. Cecere et al. (2017) used a wider range of SOAs (50-500ms), allowing for sufficiently more comparisons of onset differences between AV and VA trials which were not applied in the current experimental paradigm.

Furthermore, the sensory ERPs elicited by both auditory- and visual-first trials at the 150ms SOA, showed consistency; the only difference between the -150ms and +150ms SOA was a reduction in auditory N1 latency for A_IV_I compared to A_OV_O trials at the +150ms SOA. Although this may suggest early speeded processing of visual-first, A_IV_I events at auditory sensory sites, this finding should be interpreted with caution. Due to the prior onset of the visual stimuli eliciting positive amplitude in primary visual cortex, the negative end of the dipole was likely recorded from auditory sensory sites (FCz), resulting in initial negative amplitude prior to the onset of auditory stimuli. Thus, early visual activity at auditory sensory sites may influence proceeding ERPs and potentially the speeded latency that was observed for A_IV_I compared to A_OV_O trials. This caution of interpretation should be applied to all FCz sites at which visual stimuli onset some time prior to auditory information (+150/300ms SOAs), when interpreting auditory-evoked amplitudes and latencies (see Appendix D).

Moreover, we compared simultaneous and non-simultaneous responses by first collapsing across frequency at each SOA. Second, for an insight to frequency differences, we analyzed all four frequency conditions only at the +150ms and -300ms SOA as these included relatively equal proportion of simultaneous and non-simultaneous responses. It was highly probable that we would be unable to draw meaningful conclusions from the other three SOAs, given the large proportion of simultaneous responses for the 0ms and -

150ms SOAs, and considerably low proportion for +300ms SOA. We observed shorter latencies for visual N1 and auditory P2 components at +150ms and -150ms SOAs, respectively, but only when integration did not occur. However, we did find more sensory ERP differences for the $A_I V_I$ compared to all other frequency conditions, but only at the -300ms SOA, in bilateral suppression of visual N1 and enhancement of P2 amplitude for simultaneous compared to non-simultaneous judgements (see Figures 9 and 10). This may serve as an indicator of stronger sensory processing effects that are happening faster for integrated trials, particularly during a temporal lag that is near the limit of the TWI in adults. Specifically, we found speeded P2 latency at the PO8 electrode site but not at PO7, which only approached significance.

4.3 Limitations

Several limitations became evident as a result of the current experiment. First, our sample size (N = 21) may have contributed to marginally significant p-values, as well as a lack of some insignificant behavioral and ERP findings. Despite having substantial number of trials, the signal-to-noise ratio can be influenced by sample size, which in turn can affect ERP amplitude and latency (Luck, 2005). Additionally, this low number of subjects may have contributed to a small amount of false, “same” responses (or indications of integration) for the +/-300ms SOAs, in that, some “same” responses at these larger SOAs could have been due to participants mistakenly clicking the wrong button on the computer mouse pad.

Second, due to the accommodation for maximizing trials in ERP analyses, we decreased the number of SOAs in the current study to five, in comparison to the 13 that were originally utilized in the behavioral study by Green and colleagues (2019). Because of this, we were unable to obtain an accurate measurement of TWI width. Third, although our stimuli included frequencies that were low and high both inside and outside of the peak sensitivity range (see Appendix A), we collapsed across all inside and all outside frequencies for each modality to maximize ERP trials. Therefore, we were unable to discern potential differences between low and high frequencies inside and outside of respective peak sensitivity ranges.

4.4 Future Directions

Future studies should consider a larger recruitment of participants to avoid being potentially underpowered and mitigating the chance of possible false indications of multisensory integration. A larger sample size may also result in ERP waves that are less noisy and include an increasing number of significant differences in sensory components between frequency conditions. Additionally, future researchers should consider investigating low and high frequencies within both inside and outside of the peak sensitivity range. For example, this would allow for a clearer understanding of the impact that auditory frequencies residing inside the peak sensitivity range can have on integration and whether a difference between inside “low” and “high” exists. Similarly, frequencies outside of the peak range should be dichotomized and further examined in the context of the current experiment. Previous work has already established that faces containing low or high visual spatial frequencies are processed differently (Langner, Becker & Rinck, 2009). Faces configured with high spatial frequency are linked with recognition of identity while low spatial frequency in faces is prominent in emotional expression. Further, individuals with social anxiety disorder who already have a hyperresponsive amygdala are better at processing low spatial frequency in faces (Langner et al., 2009).

Moreover, it may be beneficial to also consider producing topographical maps to gain a better understanding of the distribution of neural activity across the scalp. In particular, conducting scalp topographies across time between integrated AV and VA trials would allow for a potential discovery of multisensory integration occurring in stages, similar to the results found by Cecere et al. (2017). Although simple, non-speech AV stimuli of varying frequencies were also included in their study, the stimuli of the current experiment contained some visual spatial and auditory sound frequencies that are much lower on the frequency spectrum. Furthermore, scalp topography distributions can offer insight to potential interactions between electrode cites and demonstrate the distributions of ERPs across the scalp (Stekelenburg & Vroomen, 2007).

Finally, future studies should consider investigating potential oscillatory effects of the current paradigm. A recent study conducted by Bastiaansen and colleagues (2020) emphasized the influence of alpha oscillatory activity on proceeding perception of stimuli. They demonstrated higher alpha oscillation prior to a subjectively simultaneous response in comparison to lower alpha activity observed prior to non-simultaneous responses. However, it is worth noting that Batiaansen et al. (2020) used simple, audiovisual stimuli only in the VA fashion, as a potential representation of natural binding of visual-then-auditory unisensory information in the environment. Future research should investigate alpha oscillation onset using simple, non-speech stimuli that vary across the peak frequency spectrum. By considering the limitations and suggestions towards future work, researchers will be able to conduct experiments to further enhance our understanding of the multisensory integration process.

4.5 Conclusion

Behaviorally, we were able to replicate nearly all major findings of Green and colleagues (2019) despite having a substantially lower number of subjects. At the neurophysiological level, we demonstrated visual ERPs that were happening faster and were overall larger for audiovisual events in which the visual stimuli were within the peak sensitivity range of human perception, such in the case for speech. This effect was consistent across each SOA condition as far as 300 milliseconds of temporal lag, which is partially in line with speech events consistently eliciting the widest TWI. Most studies have focused on effects of visual stimuli on auditory ERPs in speech-related multisensory integration (Knowland et al., 2014; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). To our knowledge, we are the first to show the criticality of visual stimuli in audiovisual events and their influence on sensory ERPs. When breaking down the frequency pairings, we found the most differences between integrated and non-integrated trials for A_IV_I trials, further supporting the uniqueness of A_IV_I events and multisensory integration.

REFERENCES

Bastiaansen, M., Berberyan, H., Stekelenburg, J. J., Schoffelen, J. M., & Vroomen, J. (2020). Are alpha oscillations instrumental in multisensory synchrony perception? *Brain Research*, *1734*, 146744.

Calvert, G., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. MIT press.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, *98*(1-3), 191-205.

Cecere, R., Gross, J., & Thut, G. (2016). Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *European Journal of Neuroscience*, *43*(12), 1561-1568.

Cecere R., Gross J., Willis A., & Thut G (2017). Being First Matters: Topographical Representational Similarity Analysis of ERP Signals Reveals Separate Networks for Audiovisual Temporal Binding Depending on the Leading Sense, *J Neurosci*, *37*(21):5274-5287. DOI:10.1523/JNEUROSCI.2926-16.2017.

Chen, Y.-C. & Spence, C. (2017). Assessing the role of the 'unity assumption' on multisensory integration: a review, *Front. Psychol.*, *8*, 445. DOI:10.3389/fpsyg.2017.00445.

De Valois, R. L., Albrecht, D. G. & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex, *Vis. Res.*, *22*, 545–559.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*(6), 719-721.

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

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, *11*(5), 473-490.

Green, J.J., Pierce, A.M., & Mac Adams, S.L. (2019). Multisensory Integration is Modulated by Auditory Sound Frequency and Visual Spatial Frequency, *Multisensory Research*, *32*, 589–611.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia*, *49*(3), 461-467.

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental science*, *15*(5), 688-696.

Holmes, P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability, *Neuropsychologia*, *45*(14), 3340-3345.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta psychologica*, *134*(3), 372-384.

Knowland, V. C., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science*, *17*(1), 110-124.

Langner, O., Becker, E. S., & Rinck, M. (2009). Social anxiety and anger identification: Bubbles reveal differential use of facial information with low spatial frequencies. *Psychological Science*, *20*(6), 666-670.

Luck, S. J. (2005). Ten simple rules for designing and interpreting ERP experiments. *Event related potentials: A methods handbook*, *4*.

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

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, *7*(10), 3215- 3229.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of neurophysiology*, *56*(3), 640-662.

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory processes: a balancing act across the lifespan. *Trends in Neurosciences*, *39*(8), 567-579.

Nelson W.T., Hettinger L.J., Cunningham J.A., Brickman B.J., Haas M.W., McKinley R.L. (1998). Effects of localized auditory information on visual target detection performance using a helmet-mounted display, *Hum Factors*, 40:452–460.

Repp, B. H. and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences, *J. Exp. Psychol. Hum. Percept. Perform*. *28*, 1085–1099.

Robinson, D. W. and Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones, *Br. J. Appl. Phys.*, *7*, 166–181.

Senkowski D., Saint-Amour D., Kelly S.P., Foxe J.J. (2007). Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study, *Neuroimage*, *36*, 877–888. DOI: https://doi.org/10.1016/j.neuroimage.2007.01.053.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, *1296*(1), 31-49.

Stein, B.E., & Standford, T.R. (2008). Multisensory integration: current issues from the perspective of the single neuron, *Nat. Rev. Neurosci*. *9*, 255–266.

Stekelenburg, J.J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events, *Journal of Cognitive Neuroscience*, *19*(12):1964-73. DOI: 10.1162/jocn.2007.19.12.1964. PMID: 17892381.

Stevenson, R.A., & Wallace, M.T. (2013) Multisensory temporal integration: task and stimulus Dependencies, *Exp Brain Res*, 227:249 –261.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology*, *13*(13), R519-R521.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech, *Proceedings of the National Academy of Sciences, U.S.A., 102*, 1181–1186.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics, 69*, 744–756. DOI: https://doi.org/10.3758/BF03193776.

Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. *Multisensory object perception in the primate brain*, 95-121.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review, *Attention, Perception, & Psychophysics*, *72*(4), 871-884.

Vroomen J., & Stekelenburg J.J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli, *J Cogn Neurosci, 22*(7):1583-96. DOI: 10.1162/jocn.2009.21308. PMID: 19583474.

Walker, J. T., & Scott, K. J. (1981). Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(6), 1327.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of neurophysiology*, *80*(2), 1006-1010.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64,* 105-123.

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). “Why sounds are judged longer than lights": Application of a model of the internal clock in humans. *The Quarterly Journal of Experimental Psychology: Section B*, *51*(2), 97-120.

APPENDIX A: EXPERIMENT STIMULI

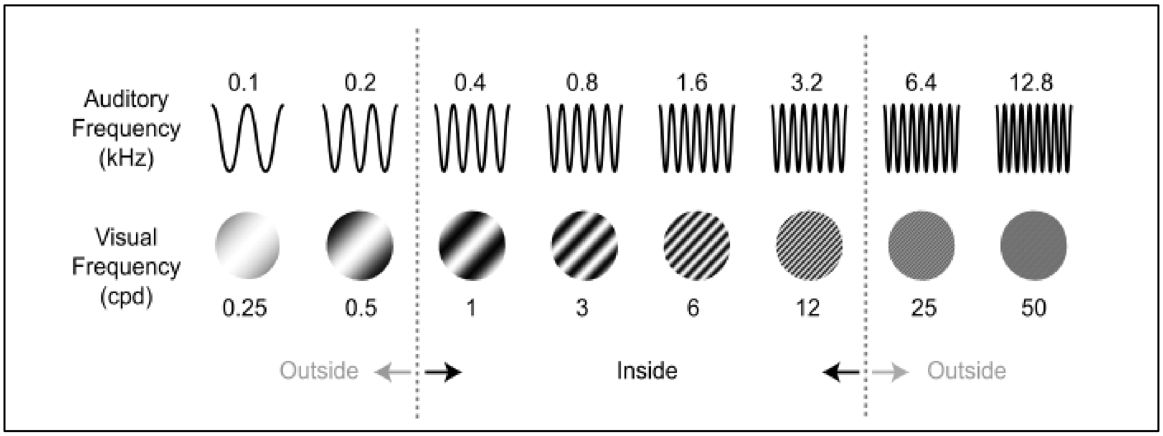
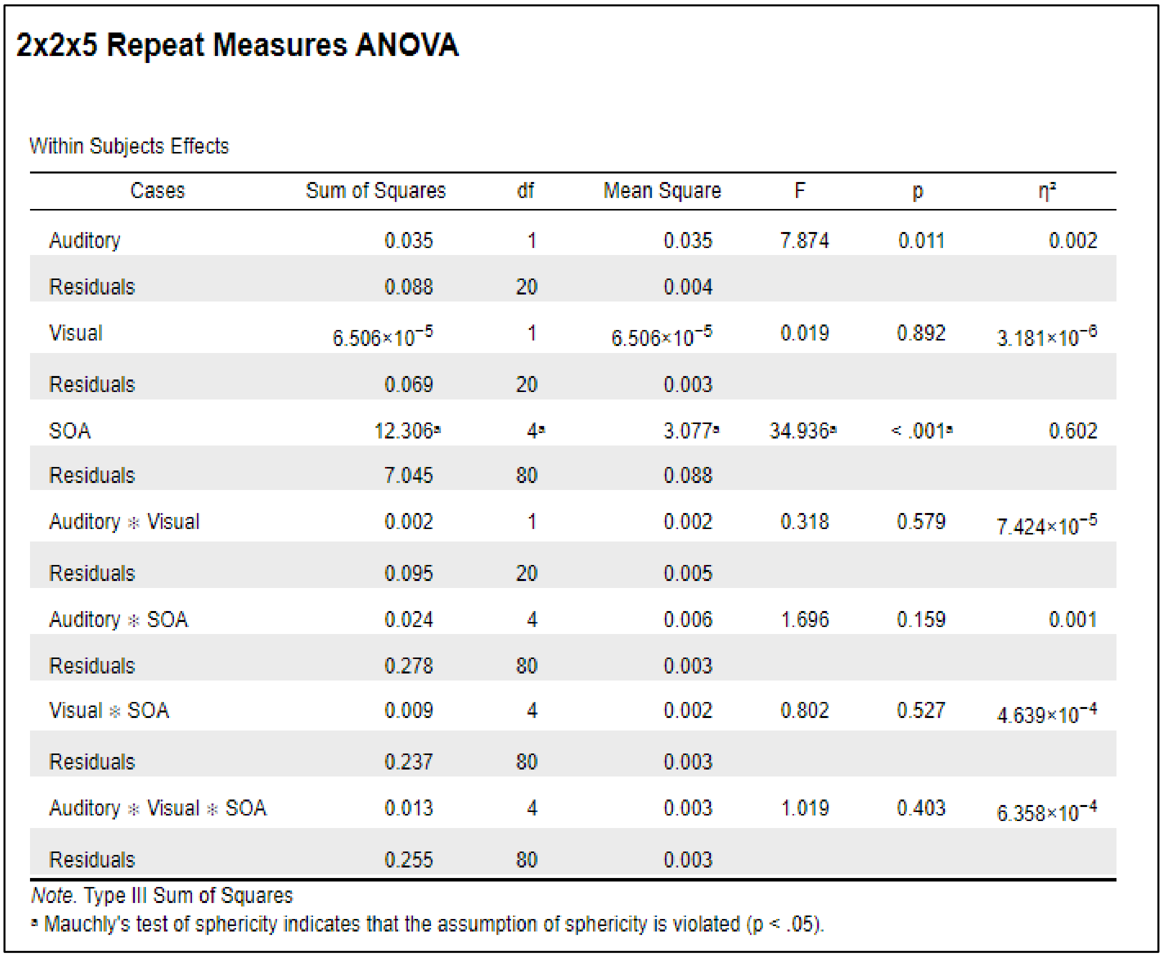


Figure A.1. Spectrum of auditory sound and visual spatial with estimated cut-offs of residing inside, or outside of, the range humans are most sensitive to.

APPENDIX B: BEHAVIORAL ANOVA REPORT



2x2x5 Repeat Measures ANOVA

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Auditory	0.035	1	0.035	7.874	0.011	0.002
Residuals	0.088	20	0.004			
Visual	6.506×10^{-5}	1	6.506×10^{-5}	0.019	0.892	3.181×10^{-6}
Residuals	0.069	20	0.003			
SOA	12.306[a]	4[a]	3.077[a]	34.936[a]	< .001[a]	0.602
Residuals	7.045	80	0.088			
Auditory ✻ Visual	0.002	1	0.002	0.318	0.579	7.424×10^{-5}
Residuals	0.095	20	0.005			
Auditory ✻ SOA	0.024	4	0.006	1.696	0.159	0.001
Residuals	0.278	80	0.003			
Visual ✻ SOA	0.009	4	0.002	0.802	0.527	4.639×10^{-4}
Residuals	0.237	80	0.003			
Auditory ✻ Visual ✻ SOA	0.013	4	0.003	1.019	0.403	6.358×10^{-4}
Residuals	0.255	80	0.003			

Note. Type III Sum of Squares

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

Table B.1. Outcomes for 2 x 2 x 5 repeated measures ANOVA on proportion of simultaneous responses.

APPENDIX C: REACTION TIME PER CONDITION

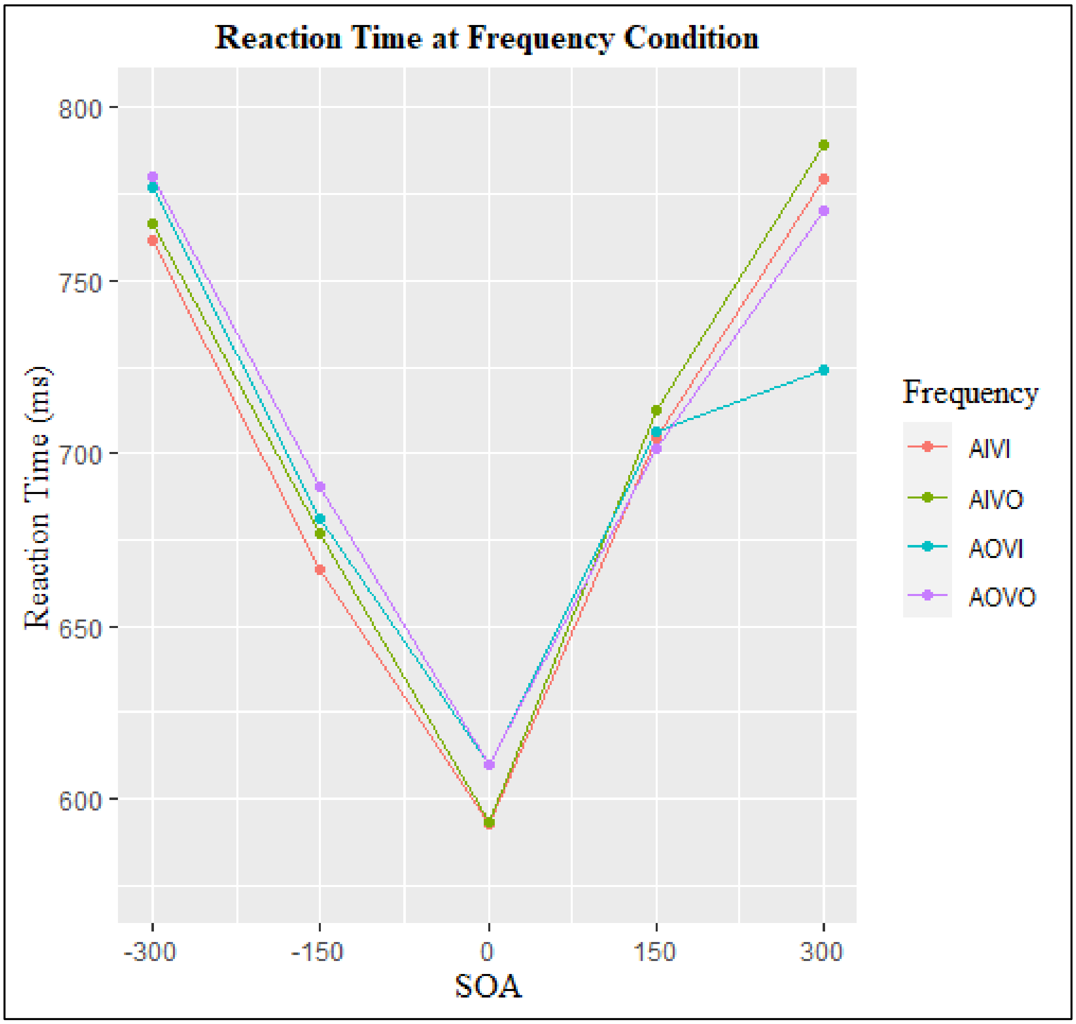


Figure C.1. Mean response time (RT) for each frequency conditions across stimulus onset asynchrony (SOA).

APPENDIX D: AUDITORY- VS. VISUAL-FIRST ERPS

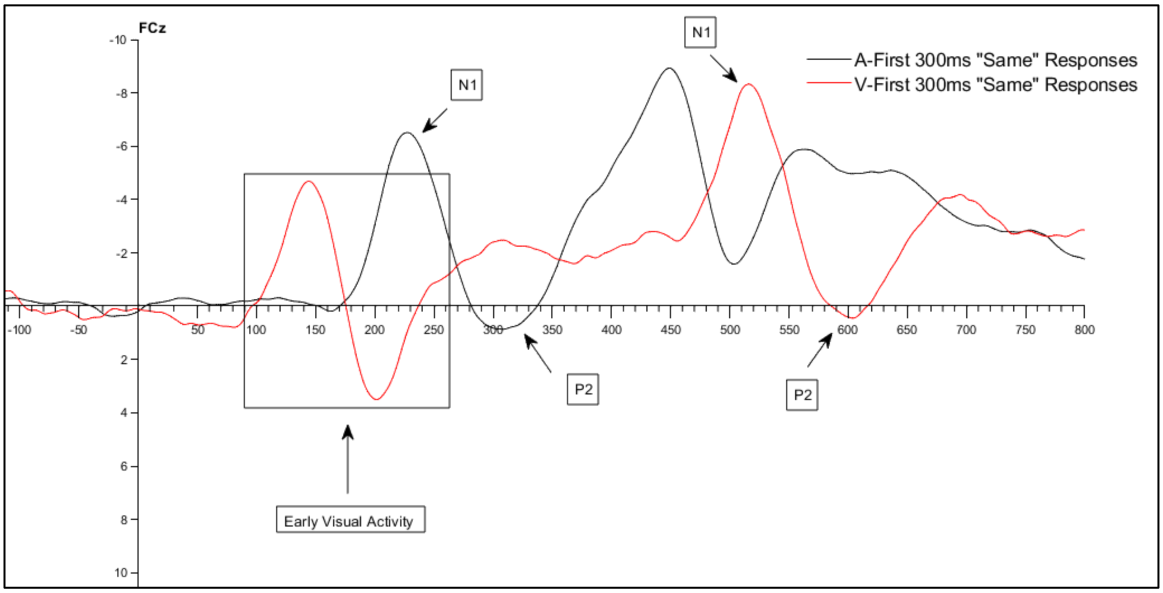


Figure D.1. ERP plot comparing auditory-first (black) and visual-first (red) events for the -300ms SOA at FCz electrode site. Mean amplitude differences between auditory- and visual-first, auditory N1/P2 were conducted and compared, with significant outcome. Early visual activity for visual-first trials can be seen around 100-250ms and potentially influencing subsequent ERPs.