



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

An ERP N400 Study: Semantic processing across modalities in the human brain.

A thesis submitted in fulfilment of the requirements for the degree

of

Master of Science in Psychology

at

The University of Waikato

by

SARAH PERRY



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2023

Abstract

Semantic processing in the brain of both language and action has been associated with the N400; an event-related potential (ERP) that is typically present when information has violated one's semantic expectations. We know that the brain receives and processes information from multiple modalities which means that cross-modal semantic processing is more aligned with how semantic processing is likely to occur in comparison to unimodal semantic processing. The N400 is a consistent effect across cross-modal language-based studies. But we know that the topographical distribution of the N400 does vary between language-based and non-language-based paradigms. Therefore, to investigate cross-modal semantic processing within the action domain, we presented participants with photographs portraying the implementation of common actions. These sequences concluded with a sound that was either congruent or incongruent with the prior action photographs. In our ERP study of 25 participants, with a mean age of 23 years (SD = 10.78 years), we¹ found an N400 effect for incongruent information processing. In addition, our findings showed a delayed N400 effect and a reduced P200 amplitude for incongruent information. These results suggest that cross-modal semantic processing of action sequences requires an increased cognitive workload which is evidenced when semantic processing does not progress as expected. Considered as a whole, these results indicate that cross-modal semantic processing is similar to unimodal processing. Cross-modal information likely requires the involvement of additional cognitive processes that are not present during unimodal paradigms.

¹ I use the word "we" here to convey that although the work in this thesis is my own, I conducted it under the guidance of my supervisor and the EEG lab technician. In other parts of this thesis I used the word "we" to convey what is or not known by the wider scientific community.

Acknowledgements

I would like to acknowledge and thank my supervisor Vincent who gave me his invaluable time and wisdom. He has introduced me to a new area and made the whole process an exciting and doable challenge, instead of a daunting task. I appreciate how you have challenged me while also respecting what is important to me.

A great thank you to Andrew who trained me in EEG, without you and your skills I may have been truly lost. I am so grateful for the time that you were able to spend with us here in New Zealand and I wish you and your family the best in the UK.

I am also grateful to the University of Waikato for the opportunity to study here and to the previous professors I have had that have prepared me well for this task; namely, Maryanne Garry. I will no doubt continue to use the skills that I learnt in the Garry Lab throughout the rest of my career.

Thank you to my participants, who took part in a rather laborious task and made data collection a really enjoyable time. And thank you to my fellow students with whom it was a joy to study alongside.

Lastly, I would like to thank my supportive husband Nathan, and my family and friends for their engagement in my project. For helpful discussions around the table, feedback that was given and many meals that were cooked for me. May all glory be to God.

Table of Contents

Abstract.....	2
Acknowledgements.....	3
Table of Contents.....	4
List of Figures.....	5
Introduction.....	7
Method.....	13
Analysis.....	20
Results.....	21
Discussion.....	23
References.....	32
Appendix A.....	41
Appendix B.....	42

List of Figures

Figure 1 <i>Presentation of visual and auditory stimuli in our cross modal paradigm.....</i>	14
Figure 2 <i>Physical layout of the experimental paradigm.....</i>	17
Figure 3 <i>Results from the Cz electrode of amplitude between Congruent and Incongruent conditions.....</i>	21
Figure 4 <i>Recorded amplitude at all electrode sites for congruent and incongruent conditions...38</i>	38
Figure 5 <i>Results from the C4 electrode of amplitude between Congruent and Incongruent conditions.....</i>	39
Figure 6 <i>Results from the CP2 electrode of amplitude between Congruent and Incongruent conditions.....</i>	39
Figure 7 <i>Results from the FC2 electrode of amplitude between Congruent and Incongruent conditions.....</i>	40
Figure 8 <i>Results from the P3 electrode of amplitude between Congruent and Incongruent conditions.....</i>	40
Figure 9 <i>Results from the CP1 electrode of amplitude between Congruent and Incongruent conditions.....</i>	41
Figure 10 <i>Results from the Pz electrode of amplitude between Congruent and Incongruent conditions.....</i>	41

Figure 11 <i>Results from the Cz electrode of amplitude between Congruent and Incongruent conditions</i>	42
Figure 12 <i>Results from the Fz electrode of amplitude between Congruent and Incongruent conditions</i>	42
Figure 13 <i>Results from the FC1 electrode of amplitude between Congruent and Incongruent conditions</i>	43
Figure 14 <i>Results from the C3 electrode of amplitude between Congruent and Incongruent conditions</i>	43

Introduction

It was discovered over 40 years ago that adults' brains produce a specific negative peak in electrical activity known as the N400 when they read the line "The pizza was too hot to cry" which is not present when they read the more logical sentence "The pizza was too hot to eat" (Kutas & Hillyard, 1980b). This work was hugely influential in the study of semantic processing, i.e. how we process information to understand and contextualise its meaning. However, semantic processing is used in many different circumstances and helps us to predict not only which word is coming next in the sentence, but also what the outcome of a situation or an action will be (Kutas & Fredermeier, 2011). Recent research has investigated the semantic processing of symbols, photographs and videos (Gunter & Bach, 2004; Kaduk et al., 2016; Reid & Striano, 2008; Reid et al., 2009; West & Holcomb, 2002). This research has so far explored the semantic processing of information presented via a single modality. We know, however, that the brain processes information from multiple modalities simultaneously (Calvert & Thesen, 2004). This study aims to understand the nature of cross-modal semantic processing in a non-linguistic-based paradigm.

Neurons in the brain communicate through chemical and electrical changes (Hajcak, Klawohn & Meyer, 2019). These electrical changes in the brain can be recorded by a sensor cap placed on the scalp which is also known as an electroencephalogram (EEG). The electrical activity recorded through an EEG reflects cognitive processes occurring in the brain. Specifically in this study, we focussed on event-related potential (ERPs) which is electrical activity in the brain that occurs in response to a specific event in the environment (Tager-Flusberg, 1999). It was through the use of EEG and ERPs in the context of semantic processing of linguistics that researchers discovered a specific neural component known as the N400 (Kutas & Hillyard,

1980a). The N400 is a negative-going peak in electrical activity that occurs around 400 milliseconds after the onset of stimuli (Kutas & Fredermeier, 2000). The N400 has tended to be used as a heuristic label to describe brain activity that occurs 200-600 ms after the onset of a stimulus and is consistent with violations of semantic expectation (Kutas & Fredermeier, 2011). Therefore the N400 is not exclusive to one particular mental operation. But we know that the N400 response is related to people's expectations when they are processing information for meaning (Reid & Striano, 2008). For example, sentences with unexpected endings such as "I drink my coffee with cream and dog" produce greater N400 responses in comparison to sentences with expected endings like "I drink my coffee with cream and sugar." (Kutas & Hillyard, 1981). To clarify, by "N400 response" I mean that a larger negative peak in electrical activity around 400 ms was observed. However, the N400 effect does not occur in response to grammatical or structural expectations such as "the pizza was too hot to EAT," word font, or other characteristics of information delivery. It is associated only with the meaning of the word and therefore can be considered an indicator of semantic processing (Kutas & Fredermeier, 2011).

Humans process various forms of information for meaning. For instance, research has found that when participants are asked to discriminate between the two photographs of faces, non-matching faces produced a larger N400 response in comparison to matching faces (Barrett & Rugg, 1990). When participants were presented with multiple photographs of related or unrelated objects, the more unrelated the photographs were, the larger the N400 response was (McPherson & Holcomb, 1999). Likewise, unrecognizable images also produce a greater N400 response in comparison to recognizable images (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999). Relatively recent research has investigated the semantic processing of meaningful hand

gestures such as a “thumbs up” in comparison to meaningless hand gestures and found that meaningless hand gestures produce a greater N400 response (Gunter & Bach, 2004). This finding is comparable to the semantic processing of words and pseudowords. Overall, these N400 responses produced from photographs, suggest that the N400 is sensitive to the semantic relationship between non-verbal stimuli. When compared to N400 components obtained from language studies, where words are presented consecutively to build a sentence, N400 responses to photographs often have a delayed latency (Reid & Striano, 2008). This implies that in comparison to "language only" paradigms, the semantic processing of photographs requires different neural systems or is a considerably more complex task.

More recently, the study of the N400 has moved beyond language and has been used to study action sequences (see Amoruso et al., 2013 for a more in-depth review). The underlying constructs present in sequential action mirror structural patterns seen in some language sequences. When both adult and infant participants are shown a video or sequence of photographs portraying someone eating with an unexpected ending, there is a greater N400 response in comparison to actions that had an expected ending (Reid & Striano, 2008, Reid et al., 2009, Kaduk et al., 2016). A similar result was found when nonverbal stories were presented to participants through sequential photographs (West & Holcomb, 2002). Stories with unexpected endings produced an N400 response. Taken together, N400 responses elicited from actions show a strong overlap with the linguistic N400 response, suggesting a potential commonality in neural generators. However, in contrast to language-based paradigms, the majority of action studies do show some variations in neural topography, with the N400 appearing to be more frontally distributed (Amoruso et al., 2013).

Given that the majority of sensory human experiences require the brain to process information from multiple modalities, it is important to study semantic processing in this context (Matusz et al., 2019). We know that the brain has the ability to integrate information from what we see, touch, hear and smell into a cohesive experience (Bremner, Lewkowicz & Spence, 2012). This ability is particularly remarkable given that the information is simultaneously received in its various forms and initially processed by specific sensory systems (Robinson & Sloutsky, 2010). For this study, we will use the terms “unimodal” and “cross-modal” to refer to behavioural tasks involving one or more sensory systems respectively as outlined by Clavert (2001). Our current understanding of unimodal semantic processing has provided a framework and foundation for how these cognitive processes may operate, but such work is ultimately artificial when contrasted with the complexities required to process information from multiple sources in everyday life. By investigating cross-modal semantic processing we hope to gain a more environmentally accurate understanding of how semantic processing occurs.

Some cross-modal paradigms have been used to study the N400 in the linguistic field. For instance, in the word pair paradigm, participants are shown two words one after the other. The target word is the second; the first is referred to as a prime word that sets the expectations for the target word. According to research, the N400 effect is greater when the target word is unrelated to the prime word compared to when a similar target word is employed (Bentin, McCarthy & Wood, 1985, Rugg 1985 & Holcomb, 1988). Using an adaptation of that word pair paradigm, researchers have demonstrated that an N400 is generated whether the words are presented visually or audibly (Holcomb & Neville 1990). Differences that were noted included the distribution of the N400 across the scalp, which varied depending on which modality the word was delivered in (Holcomb & Neville 1990). Similar effects have been found with similar

priming paradigms across modalities (Holcomb, Anderson & Grainger, 2005; Kiyonaga, Grainger, Midgley & Holcomb, 2007). Although these studies all found consistent N400 effects across modalities, the exact paradigms used do not target the sequential processing of information as seen in action sequences and the original linguistic paradigms. The action sequence paradigm requires participants to engage in semantic processing and build up an expectation about the eventual outcome of the action. When participants' expectations are violated, then we tend to see an N400 effect (Reid & Striano, 2008; Reid et al., 2009).

Some initial cross-modal work has begun to explore action sequences. In one study, participants were shown four images, taken from *Peanuts* comics (Manfredi et al., 2018). Alongside the last image of the narrative, participants heard either congruent or incongruent spoken words or sounds. Both words and sounds produced an N400 effect, but the distribution of the effect for sounds was more frontally distributed in comparison to a central-parietal distribution for words (Manfredi et al., 2018). Even though this study suggests a transferability of the N400 effect across modalities, it is focused on simultaneous semantic processing as the image and the word or sound were presented at the same time. A similar study presented an image alongside a congruent or incongruent sound to participants (Manfredi et al., 2021). Again, these researchers found an N400 effect with a parietal topography, no difference in the central area and an increased positivity for incongruent stimuli in frontal regions. Both these studies used a cross-modal approach but investigated simultaneous rather than sequential semantic processing. We know that the time given between the initial information and the target information directly influences the ERP components which may reflect different types of cognitive processing (Holcomb, Anderson & Grainger, 2005). This research was conducted using a priming paradigm which found that when the target word was delivered with a delay of

200 ms or 800 ms after the priming word there was a greater N400 response in comparison to when the target word was presented simultaneously. This research provides evidence that paradigms targeting simultaneous semantic processing require different cognitive processes than that of sequential semantic processing. Given that the N400 produced by sequential semantic processing is thought to be driven by one's expectations of what's coming next, some time between stimuli is essential for that person to develop an expectation. Overall, these studies provide evidence that the N400 response is consistent in cross-modal simultaneous semantic processing. But the transferability of the N400 and its exact characteristics for cross-modal semantic processing of action sequences are not clear due to the nature of the paradigms that have been employed thus far. For these reasons, our study will investigate the sequential semantic processing of action sequences across modalities.

The primary question in this study was: To what extent does semantic processing, as indexed by an N400 response, occur across modalities for sequential action sequences? Based on the prior literature, we hypothesised that visual action sequences that ended with incongruent sounds would produce an N400 effect in comparison to sequences that ended with congruent sounds. The secondary research question addressed in this study was: if we do find an N400 response, how does the peak amplitude, latency and topography of that N400 compare to N400 responses elicited in cross-modal and unimodal research? While the significant variation in paradigms and results of previous studies make it difficult to precisely predict topographical and temporal characteristics of the N400, based on past research we broadly anticipate that the N400 will be centralized with a slight right bias (Kutas & Fredermeier, 2011). We know that past N400s derived from action sequence research tend to have a delayed latency when contrasted with N400s produced by written sentences (Gunter & Bach, 2004; Reid & Striano, 2008). We

hypothesise that this delayed latency may be emphasized by the added complexity of our cross-modal paradigm.

A further subsidiary question was to what extent the N400 is related to other ERP components associated with attention systems in participants. In addition to the N400, the increased complexity of our task may impact the presence and behaviour of the P200 component. The P200 is a positive ERP with an expected latency of 200 ms (Sur & Sinha, 2009). The P200, which is generally distributed over the centro-frontal scalp regions, is a reaction to target auditory or visual stimuli (Sur & Sinha, 2009). The P200 can be influenced by attention and the difficulty of the task. A recent review found that across six studies an increase in cognitive workload produced a decrease in the P200 amplitude (Ghani et al., 2020). It is plausible that our cross-modal paradigm in conjunction with the complexity of violated expectations in incongruent sequences may result in a decrease in the P200 amplitude in comparison to the P200 incongruent sentences.

If our primary hypothesis is supported, it would suggest in the cognitive domain that semantics can be understood in a similar way across modalities. Further, our results would suggest that semantic processing can be measured across and between sensory modalities through the use of the N400. This work will allow us to understand semantic processing and how it occurs in a more environmentally accurate context. For the neuroscience field, it would provide a profile for how semantic processing operates across modalities and help to lay further foundations of understanding semantics in the brain.

Method

Participants

30 participants (9 identified as male, 20 identified as female and 1 identified as non-binary), with an average age of 24 years (range 18-59 years, SD of 9.05) took part in the study. All participants were students at the University of Waikato who had normal or corrected to normal vision, were not on prescription medication and, to their knowledge, had no hearing issues. In addition, participants had to meet the COVID-19 vaccination criteria laid out by the university (i.e. at least double vaccinated, with boosters recommended, and medical exemptions discussed on a case-by-case basis). Handedness was assessed via the Edinburgh Handedness Inventory (Oldfield, 1971), a questionnaire that reported right-handed dominance for all participants. Scores ranged from 73.33 to 100, with a mean of 92.95 and an SD of 7.96. For inclusion in the grand average, the data from each participant was required to have at least 50 artifact-free trials per condition. Five participants were excluded from the ERP statistical analyses because of excessive EEG artifacts. The final sample included 25 participants (17 females), with an average age of 24 years (range 18-59 years, SD of 10.78) and an average Edinburgh handedness score of 92.55 and SD of 8.03. The study was approved by the institutional ethics committee of the University of Waikato.

Apparatus and Materials

The experimental paradigm consisted of viewing stimuli presented via a computer monitor with a resolution of 1920 x 1200 pixels. Participants were asked to listen to auditory stimuli through EEG-compatible ER3C earbuds (ER3C Tubal Insert Earphones, Etymotic Research Inc.). These earphones deliver auditory stimulation through air-conducting inserted

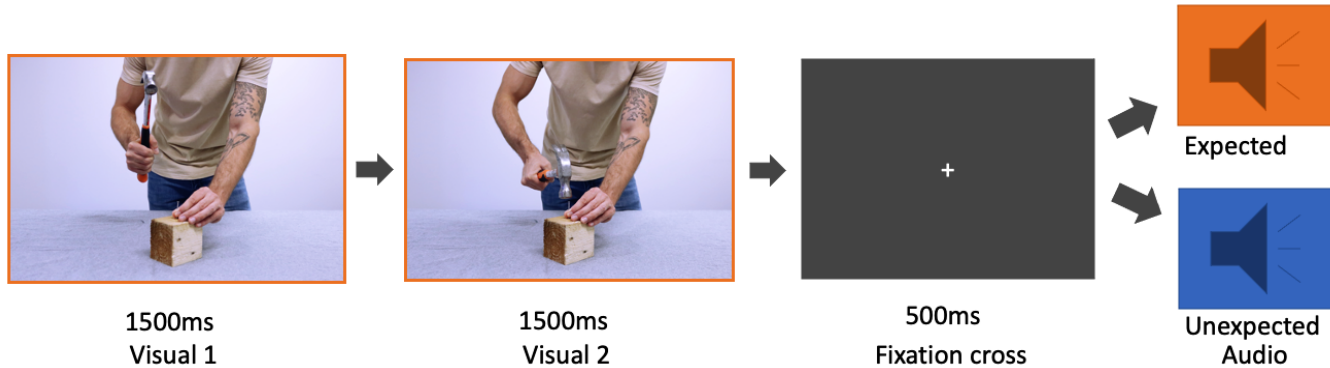
earphone tubes. These avoid any electrical artifacts being produced in the EEG due to the auditory stimulation (Boyle, 2015). These earphones have been used successfully in the field of neuroscience and have shown utility at all audiometric frequencies (Boyle, 2015; Zamm et al., 2021; Harris et al., 2021). The earphones were used in conjunction with Applied Research and Technologies (ART) Four Channel Headphone Amplifier (<https://artproaudio.com/product/headamp-4-eight-output-stereo-headphone-amp/>). The amplifier was set at zero for this study.

Visual Stimuli

The visual stimuli were modelled on those used in the comparable unimodal paradigm which is foundational to this research (Reid & Striano, 2008). We used eighteen sets of photographs depicting an actor completing a common action with a household object, such as pouring a kettle or eating an apple. There were two photographs for each activity. The first image showed the action's overall context, and the second showed the initiation of the action (see Figure 1). Each set of photographs was created to be closely matched with the only differences being central to the specific action. Out of the eighteen sets of photographs, eight had a male actor and seven had a female actor to provide some variation for the participants. An additional four sets of photographs were created and only used for the practice trials.

Figure 1

Presentation of visual and auditory stimuli in our cross-modal paradigm



Note. Participants were presented with 72 congruent (ending with an expected sound) trials and 72 incongruent (ending with an unexpected sound) in congruent trials. Trials were broken up with an interstimulus interval of 1000 ms.

Auditory Stimuli

Each set of photographs was associated with a sound (Mean duration = 1.069s, SD = 0.557s, Min duration= 0.174s, Max duration= 2.229s). For example, the photograph of the actor about to bite an apple was associated with the sound of someone biting an apple. These sounds were collected from <https://soundbible.com>, <https://freesound.org/>, and <https://www.epidemicsound.com> and edited through audio software Audacity. The sounds were edited so that they were synced to the photographs of the actions as closely as possible. For example, the crunch of an apple sound started with no delay. The sampling frequency of the sounds was also standardized at 44100 Hz.

To create congruent sequential action sequences, photographs were paired with the matching sound. For example, the photographs displaying an actor about to bite an apple would be followed by the sound of an apple being bitten. For incongruent sequential action sequences, the photographs were followed by a non-matching sound. For example, the photographs of an actor

about to eat an apple may be followed by the sound of water being poured from a kettle (see Figure 1).

To ensure incongruent action sequences were perceived as belonging to that category of actions, we asked first year psychology students at the University of Waikato to review 18 congruent and 18 incongruent sequences. Congruent was described to participants as being when “the sound matched your expectations of what you thought you would hear” and incongruent was described as when “the sound did not match your expectations of what you thought you would hear, the sound was incongruent with the photographs”. We excluded participants who did not complete the survey, did not pay attention or reported not hearing the test sound. A participant was grouped as inattentive if they wrongly classified 10 or more sequences. The final sample included 153 students. On average, congruent sequences were rated as being congruent 92.22% of the time and incongruent sequences were rated as being incongruent 96.7% of the time. When participants rated an incongruent sequence as incongruent, we asked them to rate how incongruent the sequence was on a 1 to 5 Likert scale (1= weak, 5= strong). The mean incongruency rating for incongruent sequences was 4.51 with a SD of 0.97. The data suggests that overall congruent sequences were perceived as congruent, and incongruent sequences were perceived as strongly incongruent. We can therefore be confident that the incongruent sequences used in this study did not align with the expectations of the participants.

Procedure

Prior to coming to the lab, participants completed an online survey. This survey included an information sheet explaining the procedure and what it involved along with an initial consent form. Participants also completed the Edinburgh Handedness inventory (Oldfield, 1971), and

answered questions about their hearing abilities and any prescription medication they take. Participants who were right-handed, had no hearing issues and were not on any prescription medication were then booked in for a session in the EEG lab.

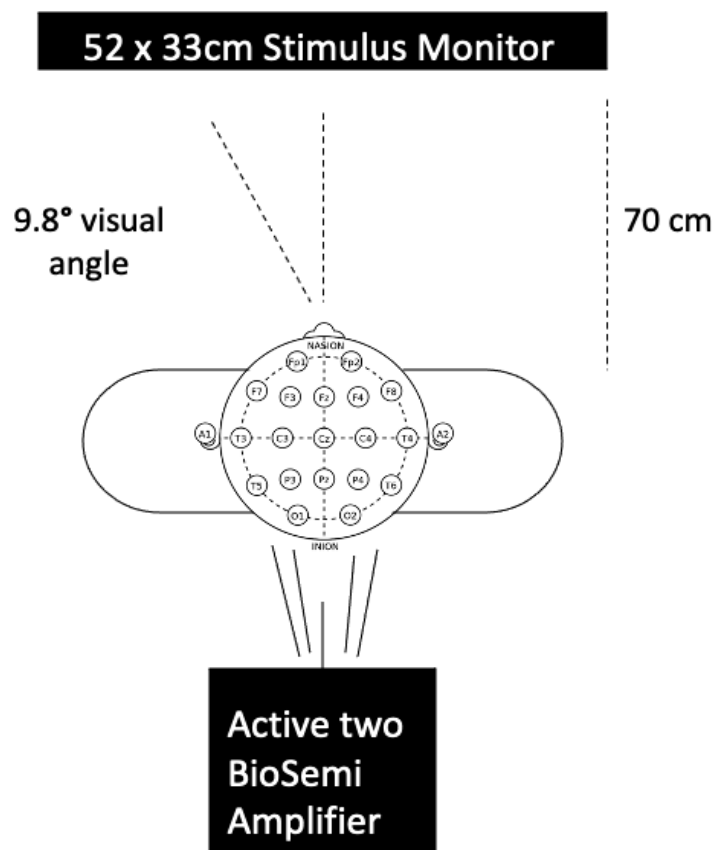
Participants were briefed verbally about what participation entailed, given a chance to ask questions and confirm their consent to participate. During face-to-face contact, mask wearing was maintained by the researchers and the participant according to institutional health and safety protocols.

A preliminary assessment of the cap's fit to the head was made after measuring the participant's head circumference to identify the appropriate size cap to employ. Six actively shielded flat electrodes for HEOG, VEOG, mastoid channels (horizontal and vertical electrooculogram) were applied to the participant with electrode gel. The Nasion electrode was left to be applied at the end. Participants were then instructed how to apply the Electrocardiogram (ECG) electrode themselves. The EEG cap was then put on the participants head with Cz as the key reference location for symmetrical positioning. Measurements were taken from the Inion to the Cz electrode and were made equal with the distance from the Nasion to the Cz electrode. Measurements were also taken from the tragus in each ear to the Cz electrode to ensure that the cap was centered. Electrode gel was then applied to each electrode hole using a blunt syringe before the electrodes were fitted into the cap to ensure good contact to the scalp. T7 and T8 electrodes were not connected until after the headphones had been fitted. Participants were instructed and assisted with fitting the in-ear headphones. The electrode contacts were then checked through the ActiView settings and the Electrode offset was determined to be below 40 mV before data collection could begin. The mean offset level across participants was 26.33 mV.

The participants were seated in a dimly lit room, free from auditory distractions. The participant sat 70 cm away from a 51.84 cm stimulus monitor with a refresh rate of 60.01Hz. The stimuli were presented at 12 by 8 cm and were thus 9.8 degrees of visual angle. The experimental paradigm is displayed in Figure 2.

Figure 2

Physical layout of the experimental paradigm



Note. The researcher sat in the corner of the room observing the participant and the data that was live streamed through to the researcher's computer. Anything of interest, such as excessive alpha waves was documented in a lab book. Apart from the Headphone amplifier that sat to the right of the monitor, the desk was clear.

Participants were provided with task instructions via the monitor. They were asked to try and keep eye and body movements to a minimum and to keep their attention focused on the stimuli presented. Participants then completed a practice run consisting of eight pseudo randomized trials, four of which were congruent and four that were incongruent. During the trials participants passively viewed the stimuli. The participants were then given a chance to ask any questions before the actual experiment began.

At the start of each trial a central fixation point was presented for 1000 ms on a grey screen. This was followed by the presentation of the initial photograph depicting an action context. This remained on the screen for 1500 ms followed by the matching action photograph which was presented for another 1500 ms. Then the screen returned to a central fixation point for 500 ms before a congruent or incongruent sound began to play. As the sounds were of various lengths, they created a natural jitter in the sequence to improve the signal-to-noise ratio of the resulting ERPs. Following the termination of the sound, the sequence would start again with a central fixation cross acting as an inter-stimulus interval (ISI) of 1000 ms. This experiment paradigm is demonstrated in Figure 1. The participants were given a short break halfway through the study.

The experiment consisted of 144 trials (72 with a congruent sound and 72 with an incongruent sound). Using PsychoPy, the subject was exposed to the two conditions in a pseudo-random order, with the restrictions that the same condition could not be presented more than three times consecutively and that the audio from trial 1 and trial 2 did not coincide, and vice versa for the visual stimuli. Every 36 trials given, there were an equal number of congruent and incongruent trials, and the same audio or visual stimuli were separated by at least 3 trials.

After completion of the EEG measurements, the cap and electrodes were removed from the participant and washed with water and detergent. All hard surfaces in the lab were also wiped down according to health and safety protocols.

EEG Recording and Analysis

Electroencephalographic (EEG) activity was recorded continuously and was time locked to the presentation of the auditory stimulus. EEG activity was recorded using 32 actively shielded Ag/AgCl electrodes within a ECI (Electro-Cap International - USA) cap and 7 actively shielded flat electrodes for the HEOG, VEOG, ECG and mastoid channels (horizontal and vertical electrooculogram and Electrocardiography). EEG data was recorded with an ActiveTwo BioSemi amplifier at a 2 kHz sampling rate with use of the inbuilt CMS/DRL loop (common mode sense and driven right leg) for signal grounding and referencing. EEG data were re-referenced offline to a linked-mastoid channel for ease of comparison with previous research that had used the same reference system.

Statistical analysis of ERP responses

EEG data was analysed using the Matlab toolbox EEGLAB vR2021b (Delorme & Makeig, 2004). First, the EEG data was filtered through a High Pass filter of 0.5Hz, a notch filter (50Hz) and a Low Pass filter of 30Hz. Inactive segments were removed from the data and HEOG and VEOG Bipolar channels were created. Ocular artefact correction was conducted using LMS (Automatic Artifact Removal extension; https://eeglab.org/others/EEGLAB_Extensions.html). The data was then separated by congruent versus incongruent trials through bin lists and the EEG recordings were segmented into epochs

that consisted of a 200 ms baseline and a 800 ms epoch from the onset of the auditory stimulus. Epochs with excessive voltages ($\pm 50 \mu\text{V}$) were rejected (Tanner, Morgan-Short & Luck, 2015 & Reid and Striano, 2008). After these epochs were rejected, five participants were excluded for having less than 50 trials per condition (Picton et al., 2000). In the remaining sample, 13.11% of all epochs were rejected. This equates to an average of 9.72 epochs in the congruent condition and 9.16 epochs in the incongruent condition per participant.

Results

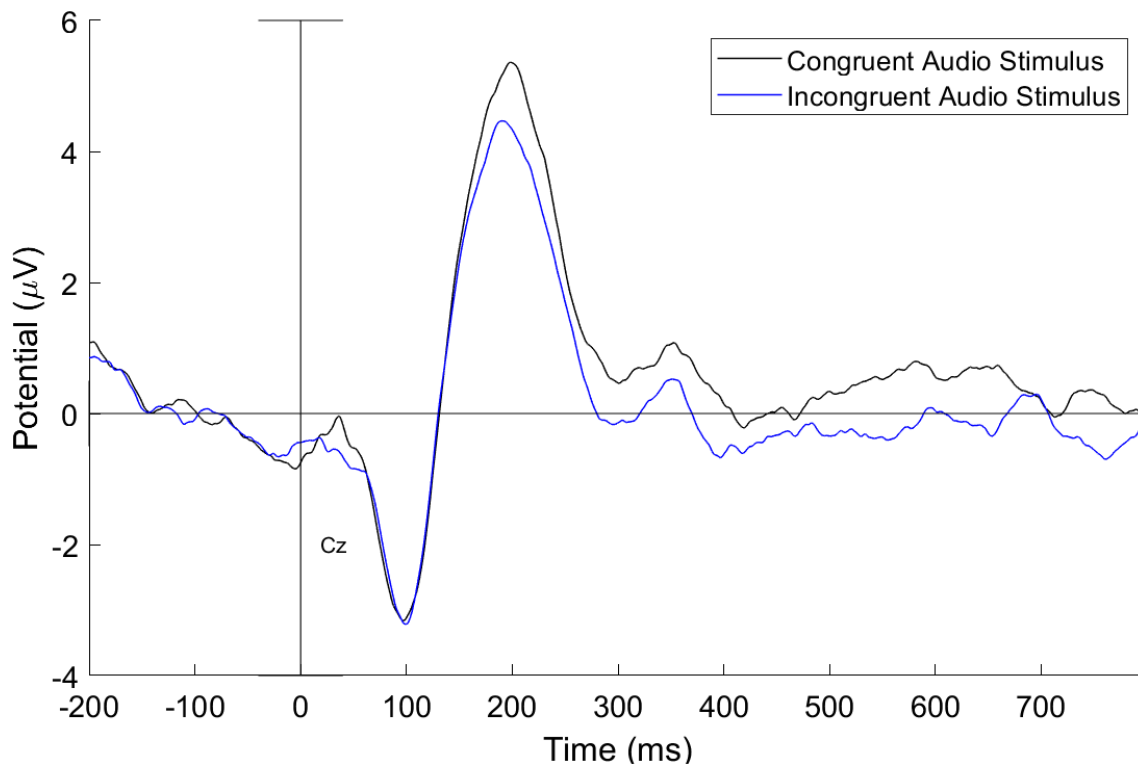
We analysed three epochs of interest: the mean amplitude voltage of the N400 in the 375-425 ms, the P200 in the 175-225 ms epoch and a Late N400 in the 525-625 ms epoch. The latency of these three epochs of interest was determined through a combination of previous research and grand average informed peak detection. For each participant we calculated the mean amplitude across each target epoch for all the trials and compared across conditions.

N400 (375-425 ms)

Based on past work we predicted that the N400 would be centralized with a slight right bias (Kutas & Fredermeier, 2011). Therefore, we compared the combined channel average of the C4, CP2, Cz, CP1, Pz, FC2, P3 electrodes between congruent and incongruent trials. We found that on average, incongruent trials produced a greater N400 effect ($M = -0.50$, $SD = 1.75$) in comparison to congruent trials ($M = 0.19$, $SD = 1.58$). This difference, 0.69, 95%CI[0.21, 1.17], was statistically significant, $t(24) = 2.99$, $p = .003$ (see Figure 3).

Figure 3

Results from the Cz electrode of amplitude between Congruent and Incongruent conditions



Note. The Cz electrode is representative of the N400 (375-425 ms), P200 (175-225ms) and Late N400 (525-625 ms) effects observed across electrodes. Note that there is no difference in baseline prior to the onset of stimulus. For further results refer to appendix A for an overview of all the electrodes and appendix B, C and D for close ups of the specific electrodes analysed for the N400, Late N400 and P200.

Late N400 (525-625 ms)

We observed a predicted downstream effect that followed the same pattern as the N400. This difference was anticipated given the complexity of the stimuli in comparison to previous studies. We examined the amplitude averages in a window from 525 to 625 ms. We compared the combined channel average of the C4, CP2, FC2, Cz, CP1, FC1, Fz electrodes between congruent and incongruent trials. We found that on average, incongruent trials produced a greater negativity ($M = -0.24$, $SD = 1.34$) in comparison to congruent trials ($M = 0.43$, $SD = 1.23$). This

difference, 0.67 , 95%CI[0.33, 1.01], was statistically significant, $t(24) = 4.07$, $p < .001$ (see Figure 5.).

P200 (175-225 ms)

We were interested in any potential impact of the added complexity of using a crossmodal paradigm. We observed a striking difference in the P200 amplitude between conditions. The P200 is known to be influenced by attention and task difficulty and it fits that this component would be influenced by the complexity of the cross-modal paradigm in conjunction with incongruent action sequences (Ghani et al., 2020). To investigate if there was a statistical difference in attention between incongruent and congruent trials, we examined differences in the P200 component across conditions. We compared the combined channel average of the Fz, FC1, FC2, C3, Cz, C4, CP1, CP2, Pz electrodes for congruent and incongruent trials. We found that on average, congruent trials produced a greater P200 amplitude ($M = 3.50$, $SD = 2.18$) in comparison to incongruent trials ($M = 2.86$, $SD = 2.08$). This difference, .65, 95%CI[.12, 1.17], was statistically significant, $t(24) = 2.53$, $p = .009$ (see Figure 4.).

Discussion

Our study investigated cross-modal semantic processing in action sequences. We showed participants an action sequence made up of two photographs followed by a congruent or incongruent sound. If information deriving from multiple modalities is processed in a similar way to unimodal information, then we expected that sequences with incongruent sounds would produce an N400 response. For sequences with congruent sounds, we predicted that there would be no N400 response. These predictions were confirmed as sequences with incongruent sounds had larger N400 responses in comparison to sequences with congruent sounds. The negativity of

the N400 peak continued into the 525 to 625 ms epoch. Finally, we found a difference in the P200 amplitude between conditions with incongruent sequences producing a decreased P200 amplitude in comparison to congruent sequences. In terms of the topography of these effects, we found that the N400 and the delayed N400 response were centralized with a slight right bias. This is a finding that is consistent with previous literature (Kutas & Fredermeier, 2011). The P200's topography was strongly centralised, which is consistent with other research showing that the P200 is typically distributed over centro-frontal scalp regions (Sur & Sinha, 2009).

These findings suggest that cross-modal semantic processing aligns with unimodal semantic processing in that they both produce an N400 response to incongruent information. The exact characteristics, however, differ depending on the specific stimulus used. For instance, the delayed N400 response is likely due to the use of visual and auditory information, given that previous visual N400s have also produced a delay (Reid & Striano, 2008). There is some variation in latency within the visual N400 studies, depending on the exact paradigm used. For example, a priming-based semantic paradigm produced an N400 response with an early latency of around 300 ms (McPherson & Holcomb, 1999). In contrast, when participants were shown short movie clips with contextually inappropriate endings, they found a delayed N400 response at 680 ms in addition to negative peaks at 315 and 550 ms (Sitnikova et al., 2008). Overall, the characteristics of our N400 response align with a consensus in the literature that suggests that the precise networks in the brain identified in any one study are heavily dependent on the exact paradigm and combination of modalities (Calvert 2001).

With the growing literature surrounding the N400 there has been an interest in localizing the source of the activity and understanding not only which areas of the brain drive this effect,

but exactly when and how they contribute (Kutas & Fredermeier, 2011). Data have been gathered using a variety of methods, one of which is intracranial monitoring of epileptic patients before surgical surgery. Patients with various types of brain injury have also provided data. With the help of these techniques and magnetoencephalographic (MEG) data, it has been determined that when it comes to linguistic semantic processing, the left temporal lobe is the major source of the N400 (See Van Petten & Luka, 2005 for a review). Although less so than the left, the right temporal lobe has also been linked to the N400 (Van Petten & Luka, 2005). More precisely, electrophysiological investigations have linked the inferior temporal regions, prefrontal regions, middle and superior temporal regions, and anterior medial temporal lobe (Kutas & Fredermeier, 2011). Other neuroimaging methods, including functional magnetic resonance imaging (fMRI) and the event-related optical signal, have also been used to find similar regions (Tse et al., 2007). We now understand that the discovered brain regions line up with the distributed network thought to be essential to the processes and storage of semantic memory (Kutas & Fredermeier, 2011). Of particular interest to this study is the neuroimaging data that points to a multimodal semantic system (Kutas & Fredermeier, 2009). For instance, MEG data obtained with scalp N400s indicates that a wave of activity takes place (Halgren et al., 1994). Starting at roughly 250 ms in the left posterior superior temporal gyrus, it spreads to the left temporal lobe by 365 ms, then between 370 and 500 ms to the right anterior temporal lobe, both frontal lobes, and other areas. According to this information, scalp N400s show a dynamic neural system that is possibly more multimodal than traditionally believed (Kutas & Fredermeier, 2011). A theory that fits given the alignment of cross-modal processing to unimodal processing as indexed by our N400.

In addition to our hypothesised findings, we found a decreased P200 amplitude for incongruent sequences in comparison to congruent sequences. Previous research has shown that

the P200 can be influenced by attention and the difficulty of the task (Ghani et al., 2020). A recent review also found that the P200 amplitude decreases in correspondence to an increase in cognitive workload (Ghani et al., 2020). So a possible explanation for this finding is that incongruent sequences required a higher level of cognitive workload in comparison to congruent sequences. Interestingly, a newly published study has also found a difference in the amplitude of the P200 between conditions. In this study, the authors manipulated the syntax and semantics of Turkish sentences and found a larger P200 amplitude for congruent conditions (Cedden, Eken, & Çakar, 2022). The authors outline cognitive workload as a possible explanation for their findings amongst other linguistic based theories. So although this study has a linguistic framework, it does provide some support for our interpretation of the P200. Our study specifically raises questions about the role of the N400 and semantic processing in domains other than language. So, it is important to note that no difference in the P200 amplitude was reported between conditions in the comparable action based unimodal paradigm (Reid & Striano, 2008). Which suggests that the P200 difference between conditions that we observed, may be driven by the combination of a cross-modal paradigm with an incongruent sequence.

There are two parts to this possible explanation of the difference in amplitude of the P200. Firstly, incongruent sequences must require more cognitive workload than congruent sequences. Secondly, cross modal semantic processing must require more cognitive workload than unimodal semantic processing. The first point that incongruent sequences must require more cognitive workload than congruent sequences does fit with our current understanding in the field of cognitive psychology. That is, when information is expected, it is easier to process in comparison to information that is unexpected therefore requiring less cognitive work (Kutas & Fredermeier, 2011). But, the second point is a bit less clear.

The link between cross-modal processing and cognitive workload is not clearly laid out in the field. But, if we consider reaction time as an index of cognitive processes and perhaps cognitive workload, then there is a collection of research we can turn to. Some research has found that when auditory and visual stimuli are presented together they are processed and responded to faster than unimodal stimuli (Molholm et al., 2002; Giard & Peronnet, 1999). This research could suggest that cross-modal processing actually requires less cognitive work than unimodal processing. This is particularly interesting as it defies the assumption that more information equals more cognitive processing. Therefore, demonstrating that crossmodal processing is not similarly the culmination of unimodal processes.

Conversely it appears that the brain responds uniquely to cross-modal information depending on whether that information fits together or not. For instance, research has found that when congruent information is presented through multiple sensory systems it is processed and responded to faster than unimodal information (Hershenson, 1962; Morrell, 1968; McGurk & MacDonald, 1976). This finding fits with neuroscience research that suggests that congruent information requires less cognitive processing because it gets bound together into a unified percept (Spence, Senkowski & Roder, 2009). Further studies have found gamma-band responses that reflect the integration of congruent information (Schneider et al., 2008; Widmann et al., 2007; Yuval-Greenberg & Deouell, 2007). However, incongruent information does not elicit any gamma-band responses. The reaction time based studies also found that when incongruent information is presented through multisensory systems it takes longer to process and respond to than unimodal information (Hershenson, 1962; Morrell, 1968; McGurk & MacDonald, 1976). If we again assume that a longer reaction time likely means that the information requires more

cognitive processing and therefore a higher cognitive load. Then, these studies suggest that crossmodal processing of congruent information has a lower cognitive workload than unimodal processing. But crossmodal processing of incongruent information, requires a greater cognitive workload than unimodal processing. This takeaway does provide some support for our possible explanation of the difference in the P200 amplitude between conditions.

Now let's consider some further implications of this study. The N400 reflects a higher-order process, which means it provides a unique opportunity to be able to study higher-level thinking and cognition, such as semantic processing. Although there has been a large interest in the N400, it can be difficult to generalize these results to real-world situations (Matusz et al., 2019). As humans we regularly encounter information presented through different sensory systems and we rely on the ability to be able to process that information to complete our everyday tasks (Robinson & Sloutsky, 2010). The crossmodal paradigm employed in this study allows us to study semantic processing, as indexed by the N400, in a context that is more closely aligned to what we experience in real-world situations. We acknowledge that this study still took place in a very controlled lab setting with lab created stimuli. So, we are by no means saying that this study is an example of naturalistic real-world research, but we believe it is a step closer in that direction. The more accurately we can understand the N400 and how it relates to different cognitive processes, the more we can use the N400 to learn about higher-level cognitive processes in general and in specific populations.

The N400 allows us to measure and understand the processing involved in typical cognitive tasks between special populations (Kutas & Fredermeier, 2011). For instance, the N400 has already been used to study and understand conditions such as Autism Spectrum Disorder,

Alzheimer's disease, Epilepsy, Schizophrenia and many more (Munte et al., 2000). The increased interest in biological reasons and mechanisms for mental diseases has also led to neuroscientific methods and findings becoming more intertwined with psychiatry and clinical psychology (Insel & Cuthbert 2015; Jones & Mendell 1999). For example, through the use of ERPs, we know that patients diagnosed with obsessive-compulsive disorder display a larger amplitude ERP component known as error-related negativity (ERN) (Gehring, Himle, & Nisenson, 2000). It is through research on the brain like our study that we can continue to build the knowledge base about the brain which can then be implemented in relevant fields such as clinical psychology.

Some limitations of our present study are that we are unable to draw any specific conclusions about where these cognitive processes are occurring in the brain. Due to the constraints of EEG research, we can be confident about the temporal accuracy of the data but we have very little spatial accuracy (Hajcak, Klawohn & Meyer, 2019). Additionally, we found our paradigm produced a significant amount of alpha waves in participants. This suggests that there were periods when participants were not actively engaging with the study (Senkowski et al., 2008). Not only does this mean that participants were most likely bored during this experiment, but these alpha waves also contribute to electrical noise in the data. The fact that we still found the results we did, despite this electrical noise, suggests that the N400 is significantly robust enough to account for additional noise in the waveform. However, future research may focus on how to make the paradigm more engaging for participants to reduce the number of alpha waves in the data.

The current study provides new insights into semantic processing in a less artificial environment than a unimodal experimental paradigm. We know that the brain is constantly given

information from multiple different sensory systems to process and understand. This work has contributed to understanding how cross-modal semantic information is manifested in the brain, by producing a cross-modal adaptation of an existing cognitive neuroscience paradigm that is more closely aligned to real-world situations. This study has shown that cross-modal semantic processing is manifested in a similar way to unimodal semantic processing. Importantly, there are some additional cognitive processes involved for cross-modal semantic processing that require a greater cognitive workload and the complexities of processing this semantic information also affects latencies of processing later in the epoch than seen in unimodal studies.

References

- Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez Gonzalez, M., Manes, F., Cardona, J. F., & Ibanez, A. (2013). N400 ERPs for actions: building meaning in context. *Frontiers in Human Neuroscience, 7*, 57.
<https://doi.org/10.3389/fnum.2013.00057>
- Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*(2), 201-212.
[https://doi.org/10.1016/0278-2626\(90\)90029-N](https://doi.org/10.1016/0278-2626(90)90029-N)
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343-355. [https://doi.org/10.1016/0013-4694\(85\)90008-2](https://doi.org/10.1016/0013-4694(85)90008-2)
- Boyle, M. (2015). Topographic characteristics of EEG responses to auditory tasks-comparisons between healthy individuals and participants with schizophrenia. <https://doi.org/10.17615/y34d-9r17>
- Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). The multisensory approach to development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 1–26). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199586059.003.0001>
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex, 11*(12), 1110-1123.
<https://doi.org/10.1093/cercor/11.12.1110>

- 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.
<https://doi.org/10.1016/j.jphysparis.2004.03.018>
- Cedden, G., Eken, A., & Çakar, T. (2022). Extracting, computing, coordination: what does a triphasic ERP pattern say about language processing?. *Language, Cognition and Neuroscience*, 37(4), 403-419.
<https://doi.org/10.1080/23273798.2021.2006726>
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9-21.
<https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological Science*, 11(1), 1-6.
<https://doi.org/10.1111/1467-9280.00206>
- Ghani, U., Signal, N., Niazi, I. K., & Taylor, D. (2020). ERP based measures of cognitive workload: A review. *Neuroscience & Biobehavioral Reviews*, 118, 18-26.
<https://doi.org/10.1016/j.neubiorev.2020.07.020>
- Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 11(5), 473-490.
<https://doi.org/10.1162/089892999563544>

- Gunter, T. C., & Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters*, *372*(1-2), 52-56.
<https://doi.org/10.1016/j.neulet.2004.09.011>
- Hajcak, G., Klawohn, J., & Meyer, A. (2019). The utility of event-related potentials in clinical psychology. *Annual Review of Clinical Psychology*, *15*(1), 71-95.
<https://doi.org/10.1146/annurev-clinpsy-050718-095457>
- Halgren, E., Baudena, P., Heit, G., Clarke, M., & Marinkovic, K. (1994). Spatio-temporal stages in face and word processing. 1. Depth recorded potentials in the human occipital and parietal lobes. *Journal of Physiology-Paris*, *88*(1), 1-50.
[https://doi.org/10.1016/0928-4257\(94\)90092-2](https://doi.org/10.1016/0928-4257(94)90092-2)
- Harris, K. C., Ahlstrom, J. B., Dias, J. W., Kerouac, L. B., McClaskey, C. M., Dubno, J. R., & Eckert, M. A. (2021). Neural presbycusis in humans inferred from age-related differences in auditory nerve function and structure. *Journal of Neuroscience*, *41*(50), 10293-10304.
<https://doi.org/10.1523/JNEUROSCI.1747-21.2021>
- Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*(3), 289–293. <https://doi.org/10.1037/h0039516>
- Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*(1), 66-85.
[https://doi.org/10.1016/0093-934X\(88\)90101-0](https://doi.org/10.1016/0093-934X(88)90101-0)

- Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*(2), 259-276. <https://doi.org/10.1006/brcg.1994.1014>
- Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281-312. <https://doi.org/10.1080/01690969008407065>
- Holcomb, P. J., Anderson, J., & Grainger, J. (2005). An electrophysiological study of cross-modal repetition priming. *Psychophysiology*, *42*(5), 493-507. <https://doi.org/10.1111/j.1469-8986.2005.00348.x>
- Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? precisely. *Science*, *348*(6234), 499-500. <https://doi.org/10.1126/science.aab2358>
- Jones, E. G., & Mendell, L. M. (1999). Assessing the decade of the brain. *Science*, *284*(5415), 739-739. <https://doi.org/10.1126/science.284.5415.739>
- Kaduk, K., Bakker, M., Juvrud, J., Gredebäck, G., Westermann, G., Lunn, J., & Reid, V. M. (2016). Semantic processing of actions at 9 months is linked to language proficiency at 9 and 18 months. *Journal of Experimental Child Psychology*, *151*, 96-108. <https://doi.org/10.1016/j.jecp.2016.02.003>
- Kiyonaga, K., Grainger, J., Midgley, K., & Holcomb, P. J. (2007). Masked cross-modal repetition priming: An event-related potential investigation. *Language and Cognitive processes*, *22*(3), 337-376. <https://doi.org/10.1080/01690960600652471>

- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463-470.
[https://doi.org/10.1016/S1364-6613\(00\)01560-6](https://doi.org/10.1016/S1364-6613(00)01560-6)
- Kutas, M., & Federmeier, K. D. (2009). N400. *Scholarpedia*, 4(10), 7790.
<https://doi.org/10.4249/scholarpedia.7790>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621. <https://doi.org/10.1146/annurev.psych.093008.131123>.
- Kutas, M., & Hillyard, S. A. (1980b, November). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language*, 11(2), 354-373. [https://doi.org/10.1016/0093-934X\(80\)90133-9](https://doi.org/10.1016/0093-934X(80)90133-9)
- Kutas, M., & Hillyard, S. A. (1980, January). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
<https://doi.org/10.1126/science.7350657>
- Kutas, M., & Hillyard, S. A. (1981). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, 11(2), 99-116.
[https://doi.org/10.1016/0301-0511\(80\)90046-0](https://doi.org/10.1016/0301-0511(80)90046-0)
- Manfredi, M., Cohn, N., Andreoli, M. D. A., & Boggio, P. S. (2018). Listening beyond seeing: Event-related potentials to audiovisual processing in visual narrative. *Brain and Language*, 185, 1-8. <https://doi.org/10.1016/j.bandl.2018.06.008>

- Manfredi, M., de Pinho, P. S. M., Marques, L. M., de Oliveira Ribeiro, B., & Boggio, P. S. (2021). Crossmodal processing of environmental sounds and everyday life actions: An ERP study. *Heliyon*, 7(9), e07937.
<https://doi.org/10.1016/j.heliyon.2021.e07937>
- Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are we ready for real-world neuroscience?. *Journal of Cognitive Neuroscience*, 31(3), 327-338.
https://doi.org/10.1162/jocn_e_01276
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748. <https://doi.org/10.1038/264746a0>
- McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, 36(1), 53-65.
<https://doi.org/10.1017/s0048577299971196>
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, 14(1), 115-128. [https://doi.org/10.1016/S0926-6410\(02\)00066-6](https://doi.org/10.1016/S0926-6410(02)00066-6)
- Morrell, L. K. (1968). Temporal Characteristics of Sensory Interaction in Choice Reaction Times. *Journal of Experimental Psychology*, 77(1), 14-18.
<https://doi.org/10.1037/h0021289>
- Münte, T. F., Urbach, T. P., Düzel, E., & Kutas, M. (2000). Event-related brain potentials in the study of human cognition and neuropsychology. In F. Boller, J. Grafman, &

G. Rizzolatti (Eds.), *Handbook of neuropsychology: Sect 1: Introduction, Sect 2: Attention* (pp. 139–235). Elsevier Science Publishers B.V..

Oldfield, R. C. (1971). *Edinburgh Handedness Inventory*. APA PsycTests.

<https://doi.org/10.1037/t23111-000>

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*(2), 127-152.

<https://doi.org/10.1111/1469-8986.3720127>

Reid, V. M., & Striano, T. (2008). N400 involvement in the processing of action sequences. *Neuroscience Letters*, *433*(2), 93-97.

<https://doi.org/10.1016/j.neulet.2007.12.066>

Reid, V. M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: evidence for semantic processing systems. *Developmental Psychology*, *45*(3), 620.

<https://doi.org/10.1037/a0015209>

Robinson, C. W., & Sloutsky, V. M. (2010). Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 135-141.

<https://doi.org/10.1002/wcs.12>

- Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22(6), 642-647.
<https://doi.org/10.1111/j.1469-8986.1985.tb01661.x>
- Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage*, 42(3), 1244-1254.
<https://doi.org/10.1016/j.neuroimage.2008.05.033>
- Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, 20(11), 2037-2057.
<https://doi.org/10.1162/jocn.2008.20143>
- Spence, C., Senkowski, D., & Röder, B. (2009). Crossmodal processing. *Experimental Brain Research*, 198(2-3), 107–111. <https://doi.org/10.1007/s00221-009-1973-4>
- Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, 18(1), 70. <https://doi.org/10.4103/0972-6748.57865>
- Tager-Flusberg, H. (Ed.). (1999). *Neurodevelopmental disorders*. MIT press.
- Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, 52(8), 997-1009.
<https://doi.org/10.1111/psyp.12437>

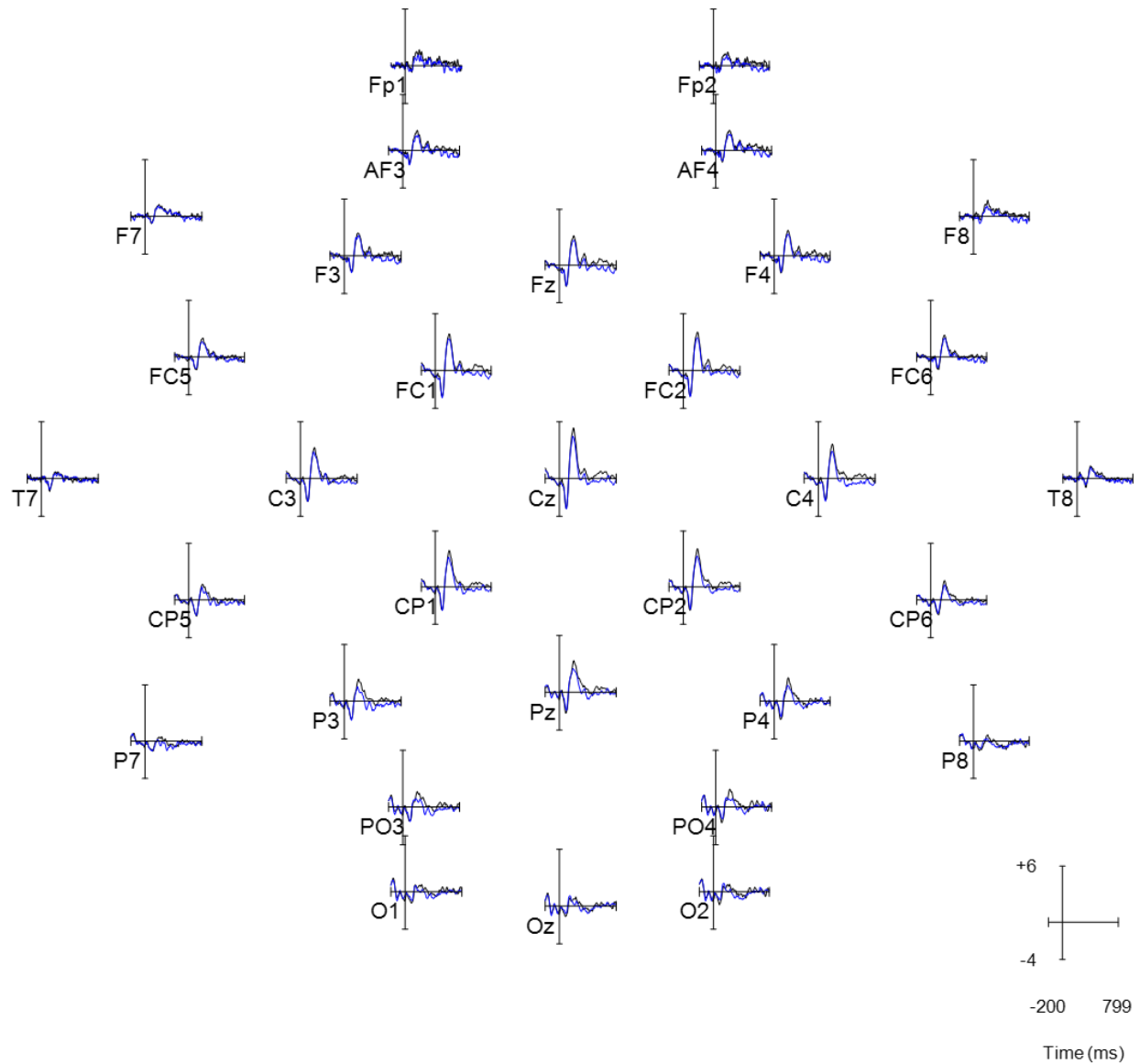
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76-82. <https://doi.org/10.1126/science.1135935>
- Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*(3), 279-293. <https://doi.org/10.1016/j.bandl.2005.11.003>
- Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*(4), 485-508. [https://doi.org/10.1016/0028-3932\(94\)00133-A](https://doi.org/10.1016/0028-3932(94)00133-A)
- West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*(3), 363-375. [https://doi.org/10.1016/S0926-6410\(01\)00129-X](https://doi.org/10.1016/S0926-6410(01)00129-X)
- Widmann, A., Gruber, T., Kujala, T., Tervaniemi, M., & Schröger, E. (2007). Binding symbols and sounds: evidence from event-related oscillatory gamma-band activity. *Cerebral Cortex*, *17*(11), 2696-2702. <https://doi.org/10.1093/cercor/bhl178>
- Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience*, *27*(5), 1090-1096. <https://doi.org/10.1523/JNEUROSCI.4828-06.2007>

Zamm, A., Debener, S., Konvalinka, I., Sebanz, N., & Knoblich, G. (2021). The sound of silence: an EEG study of how musicians' time pauses in individual and joint music performance. *Social Cognitive and Affective Neuroscience*, *16*(1-2), 31-42.
<https://doi.org/10.1093/scan/nsaa096>

Appendix A

Figure 4

Recorded amplitude at all electrode sites for congruent and incongruent conditions



Note. Firstly, the blue line refers to incongruent trials and the black line refers to congruent trials. For the 315-475 ms epoch we analysed the C4, CP2, Cz, CP1, Pz, FC2, P3 electrodes. For the 525-625 ms epoch we analysed the C4, CP2, FC2, Cz, CP1, FC1, Fz electrodes. For the 175-225 ms epoch we analysed the Fz, FC1, FC2, C3, Cz, C4, CP1, CP2, Pz electrodes.

Appendix B

Figure 5

Results from the C4 electrode of amplitude between Congruent and Incongruent conditions

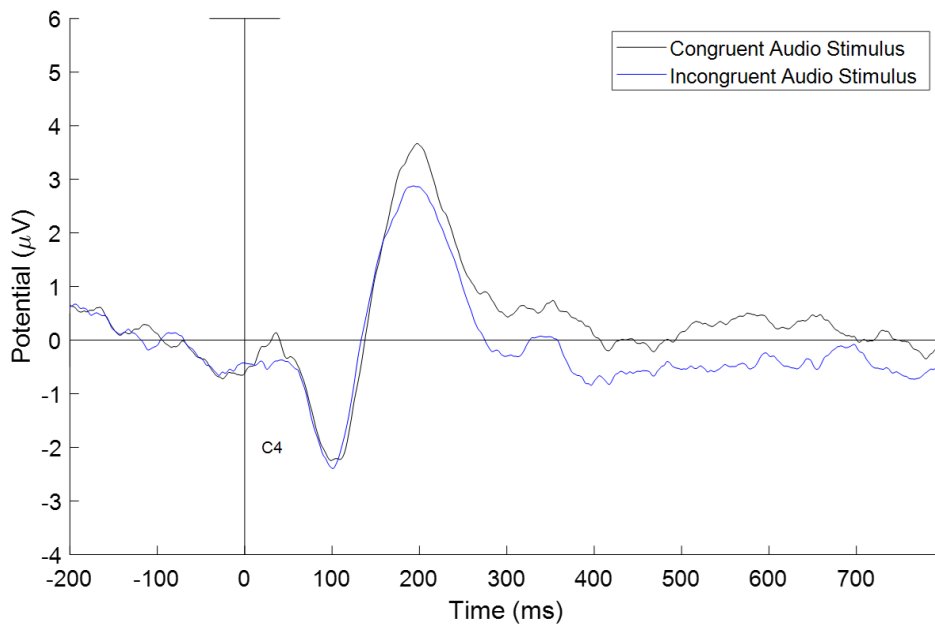


Figure 6

Results from the CP2 electrode of amplitude between Congruent and Incongruent conditions

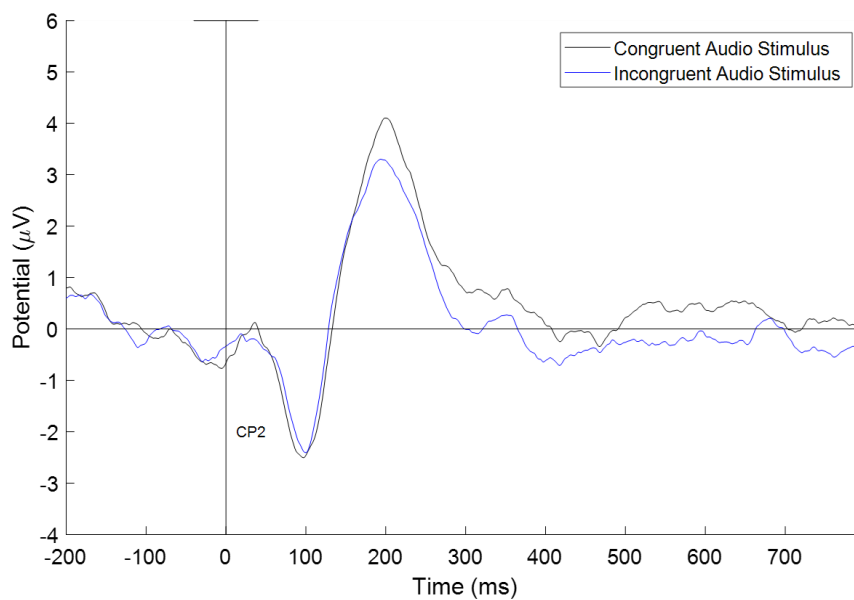
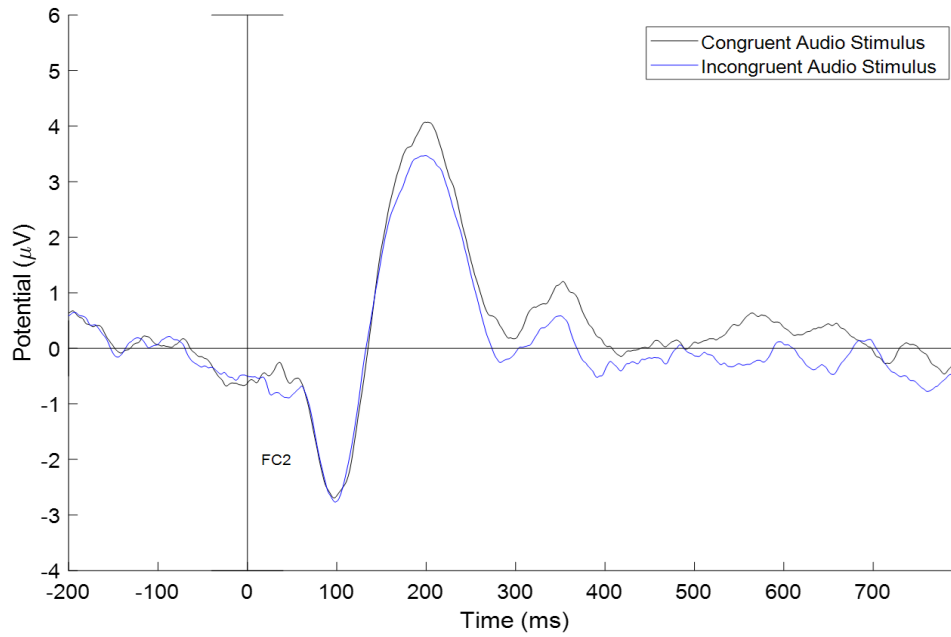


Figure 7

Results from the FC2 electrode of amplitude between Congruent and Incongruent conditions

**Figure 8**

Results from the P3 electrode of amplitude between Congruent and Incongruent conditions

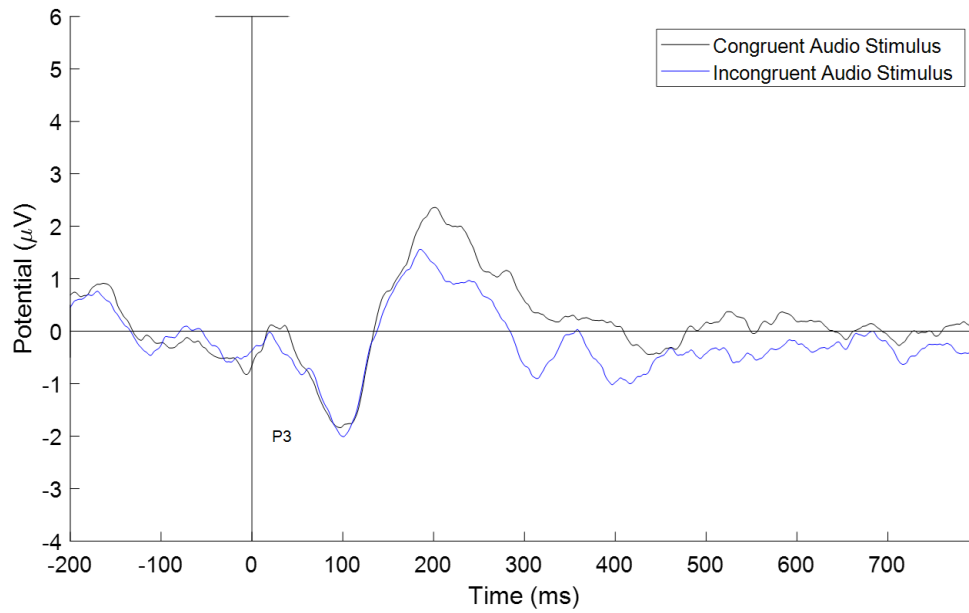
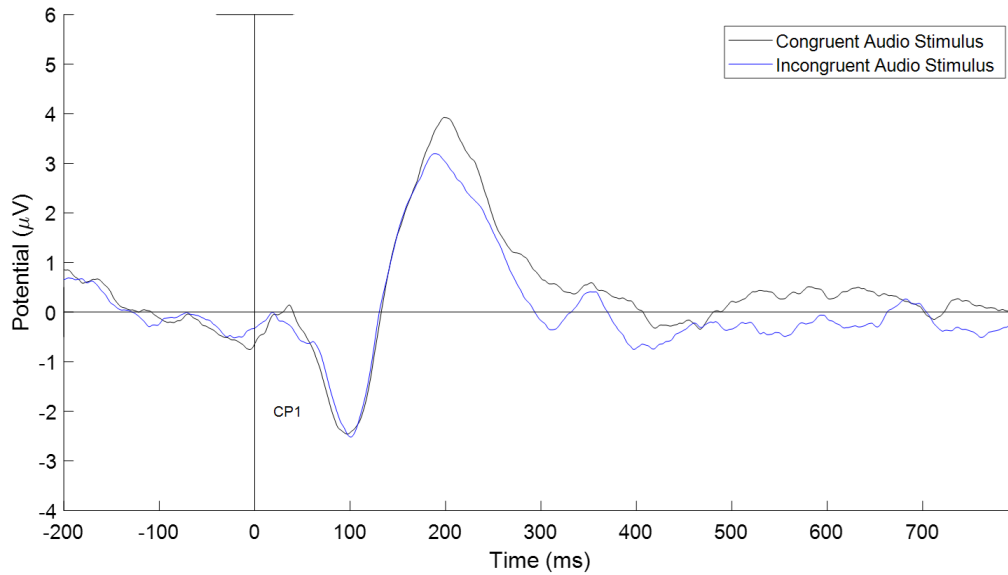


Figure 9

Results from the CP1 electrode of amplitude between Congruent and Incongruent conditions

**Figure 10**

Results from the Pz electrode of amplitude between Congruent and Incongruent conditions

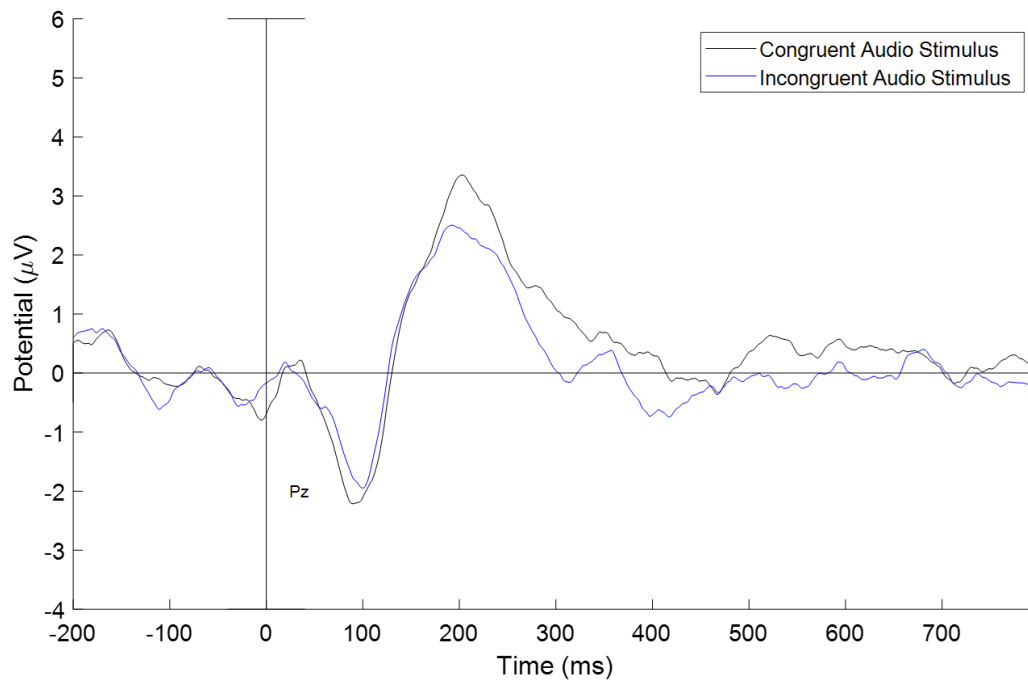
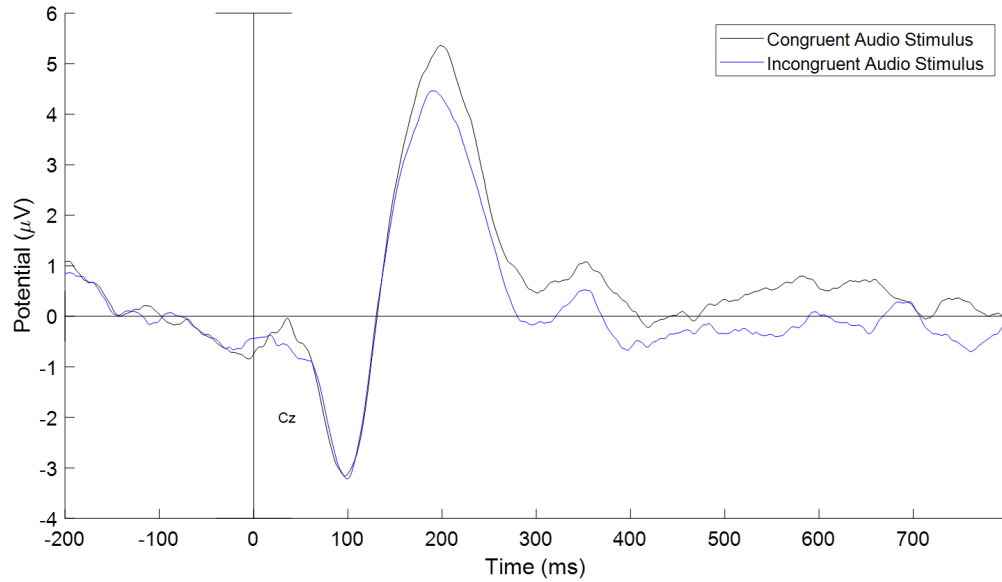


Figure 11

Results from the Cz electrode of amplitude between Congruent and Incongruent conditions

**Figure 12**

Results from the Fz electrode of amplitude between Congruent and Incongruent conditions

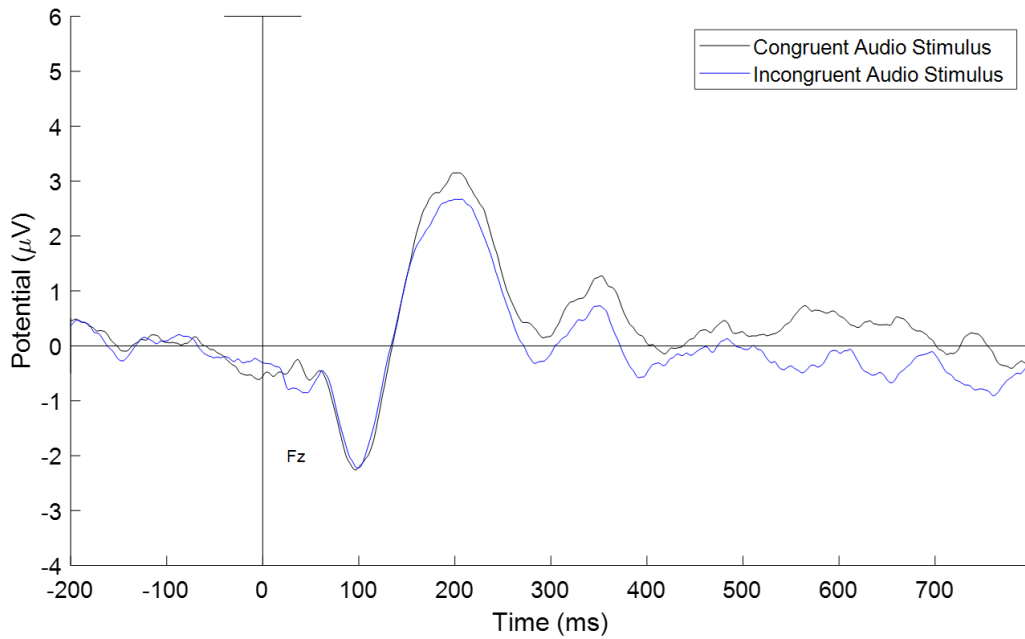
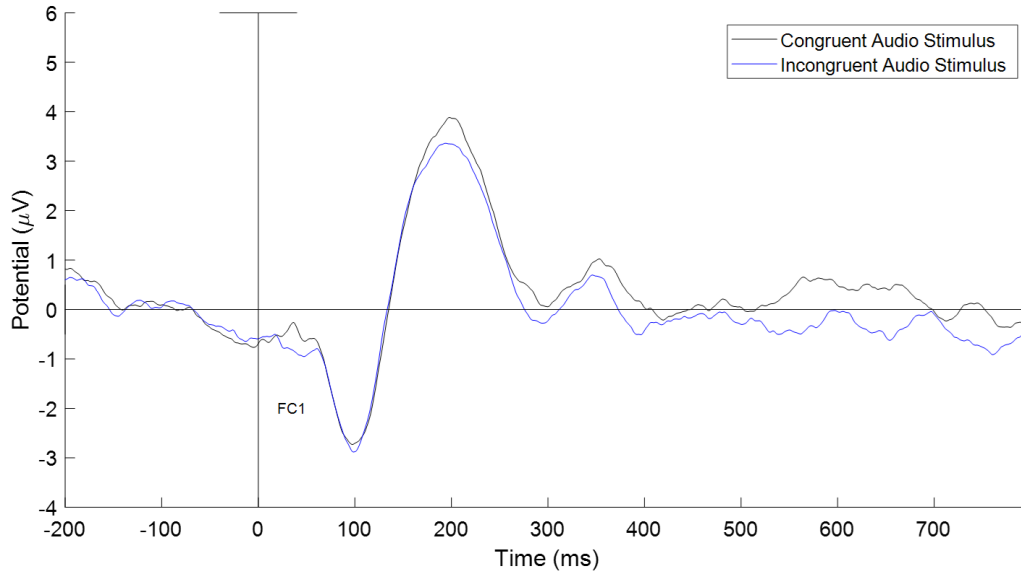


Figure 13

Results from the FC1 electrode of amplitude between Congruent and Incongruent conditions

**Figure 14**

Results from the C3 electrode of amplitude between Congruent and Incongruent conditions

