Neural Correlates of Crossmodal Correspondence Between Pitch and Visual Motion

Student ID: 33539541

Supervisor: Prof Joydeep Bhattacharya FRSA

Department of Psychology

Goldsmiths, University of London

August 2018

Thesis submitted in partial fulfilment of an MSc in Music, Mind and Brain

Table of Contents

Abstract

Crossmodal correspondences play an essential role in the construction of our realities, allowing us to efficiently process incoming multisensory information and make sense of our surroundings. In the present study, a mixed methods behavioural and EEG study inspired by Maeda, Kanai, and Shimojo (2004) exploring a crossmodal correspondence between auditory stimuli and visual motion perception was conducted. The aim was to investigate the neural underpinnings of the reported auditory-visual illusion and to better understand the nature of the crossmodal correspondence. Participants were exposed to an ambiguously moving Gabor patch and a series of ascending/descending pitches or the spoken words "up" and "down" in English and Japanese, at five different stimulus onset asynchronies, and were asked to indicate the direction of the perceived visual motion. Behavioural and ERP analyses were completed and statistically tested. It was found that both pitch content and semantic meaning of auditory stimuli are enough to bias our perceptions of visual motion, and these results help to shed light on how similar crossmodal correspondences may operate.

Introduction

Often, we associate changing pitches with moving objects—two entities which are not necessarily linked. For example, a swimmer cliff-diving into the ocean may be paired with a descending pitch, whilst the same person jumping upwards on a diving board may be paired with an ascending pitch. One of the most well-studied examples of this is the so-called "ventriloquist effect", whereby for auditory and visual stimuli presented simultaneously but at different places, the auditory stimuli are perceived to arise from a position closer to the visual stimuli as the two modalities are combined (Bertelson & Radeau, 1981; Jack & Thurlow, 1973). In fact, these seemingly arbitrary patterns of connections extend all the way through our five senses, with this effect being observed between nearly all different pairs of sensory modalities. Whereas the present study is concerned with how changing pitches may induce alternate perceptions of visual motion, there is, however, a plethora of research documenting associations across different modalities, such as audition and touch (Walker & Smith, 1985; Yau et al., 2009), audition and smell (Belkin et al., 1997; Hornbostel, 1931), vision and touch (Ludwig & Simner, 2013, Martino & Marks, 2000), and audition and taste (Bronner et al., 2008; Crisinel & Spence, 2010) to name a few. As stimuli originating from different sources converge at a neural level, we use multisensory integration to differentiate between them—enabling us to decide which stimuli should be considered a single percept and which should be segregated (Bien, ten Oever, Goebel, & Sack, 2012). In addition to unisensory neurons, the brain contains multisensory neurons which have receptive fields stemming from more than one modality, and whose purpose is to integrate the different sensory inputs if they are sufficiently spatially and temporally congruent (Wallace & Stein, 1996). Multisensory neurons have been found in both the superior temporal cortex and in the parietal cortex (Calvert, 2001; Calvert, Campbell, & Brammer, 2000; van Atteveldt, Formisano, Goebel, & Blomert, 2004). These so-called crossmodal correspondences hold an essential role in forming the basis of our reality, allowing us to process the incoming multisensory information and make sense of the world around us.

Let us begin by defining a few terms. The term *crossmodal correspondence* is used to refer to a pairing between dimensions of a stimulus or event across different sensory modalities (Spence, 2011). They occur such that an extreme stimulus in one modality should be paired with a similar extreme in the other modality. Additionally, the crossmodal correspondence is not unique to just a certain subset of the population, but instead is common for most, if not all. The stimuli can be related in many ways, from a basic to a more complex level. Firstly, an amodal dimension of the stimulus such as its temporal occurrence can be shared between modalities to create a crossmodal correspondence (Marks, Szczesiul, & Ohlott, 1986). However, they can also be related on a higher level, for example when stimuli share a quality of pleasantness, share a semantic meaning, or affect arousal in a similar way. The modulation of *multisensory integration* has traditionally been the most popular method for studying crossmodal correspondence and how the brain decides which stimuli to combine and which to separate. Besides *spatiotemporal congruency* (the overlap of two stimuli in time and space), the two most important factors affecting multisensory integration are *semantic* and *synesthetic congruency*. The former refers to the extent to which the pairs of stimuli overlap in terms of their meaning, e.g., a picture of a lightning bolt and the sound of a lightning strike. Alternatively, the latter refers to the correspondences between more basic physical features such as brightness, pitch, or size (Büttner, 2017).

Crossmodal correspondences are believed to arise due to a few different reasons. Firstly, they may come about due to the stimulus pair being naturally correlated, such as the universal link between the mass of an object and its resonant frequency—larger objects possess lower resonant frequencies and, therefore, large visual stimuli may be naturally paired with low pitches (Coward & Stevens, 2004; Grassi, 2005). Secondly, they may arise because of the way that neural connections are organised in the perceptual system. For example, magnitude-related dimensions may naturally be combined due to the fact that we represent magnitude in an identical way throughout the brain, according to the theory proposed by Walsh (2003). Thirdly, they may come about when the phrases that people use to describe one stimuli overlap with the phrases that people use to describe another. For example, the word "rough", which is used to describe both the visual or tactile quality of an object and the timbral features of a sound (Gallace & Spence, 2006; Martino & Marks, 1999). These three types of crossmodal correspondences have been termed *statistical*, *structural*, and *semantically mediated* respectively, and are thought to influence whether the correspondence is more perceptual or decisional in nature.

There is no current consensus on the true nature of crossmodal correspondences—whether they occur at a perceptual or decisional level. As for the perceptual versus decisional debate, there appears to be evidence for both. Studies by Kitagawa and Ichihara (2002) investigating perception of auditory and visual motion and by Smith, Grabowecky, and Suzuki (2007) investigating perception of androgynous faces both suggest that crossmodal correspondences possess the ability to alter participants' perceptions. Kitagawa and Ichihara (2002) showed that humans' perception of auditory intensity can be affected by exposure to visual motion in depth. Whereas many examples of a visual-motion aftereffect have been documented, where continued exposure to a changing image makes a stationary image appear to move (e.g., Anstis, Verstraten, & Mather, 1998; Tootell et al., 1995), their research shows that a combined visual and auditory aftereffect is possible. They exposed participants to a square moving in depth for a few minutes before presenting a steady sound. What they observed was that participants perceived the steady sound as changing in loudness, an example of an auditory aftereffect occurring because of a changing visual stimulus. A similar study showing the how crossmodal correspondence can affect perception is that of Smith et al. (2007). Here, the role of visual-auditory integration in perception of objects was investigated using pure tones and face gender judgements. Androgynous faces were shown alongside pure tones, and it was observed that participants more often judged the face as male when the pure tone was in the male fundamental speaking frequency range and vice versa for female judgements. Moreover, in a study by Parise and Spence (2009), it was hypothesised that if audio-visual crossmodal correspondences are truly occurring at a perceptual level, then for a pair of auditory and visual stimuli presented closely in time, participants should find it harder to determine which had been presented first on crossmodally congruent as opposed to crossmodally incongruent trials—a result that they ended up observing.

However, contrasting evidence has been found for crossmodal correspondence occurring on a more decisional level (e.g., Gallace & Spence, 2006; Marks, Ben-Artzi, & Lakatos, 2003). Participants performed a speeded visual size discrimination test in the study by Gallace and Spence (2006), whilst low/high frequency sounds or the words “high” or “low” were presented. The results showed that the congruency/incongruency of the sounds had an impact on the speed at which participants judged the size of the visual stimuli. Furthermore, the authors suggest that it may occur at a semantic rather than perceptual level, using a method where perceptual information is first converted into a linguistic code, and *then* used to help in the classification task. Similarly, Marks et al. (2003) . The most parsimonious explanation seems to be that crossmodal correspondence can occur on *both* decisional and perceptual levels, and the nature depends on the stimuli and modalities in question. These behavioural studies highlight how combining aspects of two modalities might affect judgements on a perceptual or decisional level, however, the literature is lacking brain imaging evidence for these effects—something that the present study hopes to contribute to.

On the topic of the true nature of crossmodal correspondences, another issue worth mentioning is the question of whether they are simply a normal part of a synesthetic spectrum, or can be explained in terms of Bayesian priors. It has been argued that crossmodal correspondences are similar and may in fact share the same neural mechanisms as synaesthesia (Martino & Marks, 2001; Sagiv & Ward, 2006; Ward, Huckstep, & Tsakanikos, 2006). It follows that we may all lie along a spectrum from non-synesthetic to fully synthetic behaviour. However, if synesthetes and non-synesthetes were in fact alike, it may be expected to observe enhanced multisensory integration when synesthetes are presented with crossmodally congruent stimuli, as compared to non-synesthetes (Spence, 2011), and it appears that this is not the case (Gallace & Spence, 2006). Another way to think about such correspondences is that they are purely statistical in nature and may arise from repeated exposure to the natural environment—that is, they can be explained in terms of Bayesian priors (Ernst, 2006; Ernst & Bülthoff, 2004). The idea is that stimuli may be combined based on prior knowledge (Bayesian priors) and the statistical likeliness of them being paired together. The stronger the coupling, the more probable it is that the two separate stimuli will be merged into a single multisensory percept (Spence, 2011). Therefore, crossmodal correspondences based on naturally occurring statistical correlations helps the brain combine stimuli from separate modalities. This probabilistic view of the emergence of crossmodal correspondences is unlike the synesthetic explanation, which suggests instead that additional or a lack of neural connections between neurons of the brain coding for different modalities may modulate how we integrate multisensory information. Perhaps additional replications of crossmodal behavioural studies using brain imaging methods may shed light on whether it is an affect mediated by statistical or structural differences.

The present study is inspired and based upon a paper by Maeda, Kanai, and Shimojo (2004) that reports on an illusion in which auditory stimuli alter visual motion perception in a sample of 12 participants (4 female). Their behavioural experiment consisted of two superimposed, oppositely moving Gabor patches presented alongside ascending/descending/broadband pitches and the words “up” and “down” spoken in Japanese. In one additional experiment, the stimulus onset asynchrony (SOA) between visual and auditory stimuli was varied. Participants made a two-alternative forced choice based on which direction they perceived the (ambiguous) visual stimuli to be moving, and for the words experiment they were split into groups consisting of Japanese and non-Japanese speakers. Despite the fact that the visual stimuli were never moving in either direction, it was found that participants more frequently perceived visual stimuli in the congruent direction (i.e. they saw upwards motion when the sound was ascending). The maximum effect was observed when the auditory stimuli onset is slightly after the visual stimuli onset but still completely overlapping, which the authors suggest hints at the fact that the illusion occurs at a *perceptual* level. However, they found no significant group differences between Japanese and non-Japanese speakers for the experiment involving Japanese words as the auditory stimuli. Moreover, the maximal effect was found when the voices were presented 400 ms after the Gabor patches. This suggests that this crossmodal correspondence is not completely semantically mediated, and the words are influencing the decisional and not perceptual level. Additionally, the authors completed an eye tracking experiment to show that eye-movement could not be a confound. They also altered the direction of motion of the visual stimuli to check that the effect was only present for upwards/downwards motion and to confirm that it is truly a perceptual crossmodal correspondence between motion and pitch direction.

The present study aims to replicate the behavioural aspect of the experiments by Maeda et al. (2004), whilst adding an EEG component to delve into the neural underpinnings of this changing pitch induced visual motion illusion. By including EEG analysis, we hope to understand better what is happening in the brain during this particular crossmodal correspondence, and perhaps understand the exact nature of this illusion and if it is occurring on a perceptual or decisional level. We present two experiments, testing how pitch (over five different SOAs) and English/Japanese words (presented simultaneously) affect motion perception. We hypothesise that participants will be biased towards choosing the direction of visual motion that is congruent with the sound direction presented, and that the maximal effect will be observed close to the 0 ms SOA due to the perceptual nature of the reported illusion. This is predicted due to the previous findings of Maeda et al. (2004) and the fact that if this is a semantically mediated crossmodal correspondence, then upwards pitch information should translate to upwards visual motion information and semantically congruent word stimuli should elicit the same effect. This paper will begin with a discussion of the methods used, including a detailed description of the participants and stimuli used, as well as a summary of the experimental and analytical procedures carried out. Next, the results of the behavioural and EEG experiments will be presented, before finally a discussion and interpretation of the reported findings is given, with notes on ideas for future research.

Method

Participants

The participants consisted of an opportunity sample of 31 healthy individuals aged between 20 and 51 years old (*M* = 26.7, *SD* = 6.2; 23 females). All participants were fluent English speakers (self-reported), however, for 18, English was not their native language. No individuals possessed Japanese language abilities, validating the use of the control Japanese words in Experiment 2 (see Procedure section). Additionally, participants had normal or corrected-to-normal vision, normal hearing, and all except one were right-handed (self-reported). Participants gave written consent to take part in the EEG study, which had been approved by the Internal Ethics Committee at Goldsmiths College, University of London. They received a monetary compensation of £10 for their participation in the experiments. Upon completing all tasks, they were given a full written and verbal debrief (see Appendix A1 and Appendix A2 for consent and debrief forms). One participant was omitted due to incomplete EEG recording during the task, and several others failed to pass the artefact rejection cut-off and were omitted due to noisy data. Therefore, the final data set consisted of 23 healthy individuals aged between 20 and 36 years old (*M* = 25.9, *SD* = 4.3; 15 females), 15 of which did not have English as their native language.

Materials and Stimuli

The study consisted of 2 experiments—one focussing on pitch and the other on speech. Both experiments contained auditory and visual stimuli. In Experiment 1, there were three different auditory stimuli: ascending and descending pure tones changing from 0.3 to 2.0 kHz and 2.0 to 0.3 kHz respectively, and broadband noise containing the same pitch information as the ascending/descending tones. All pitch-based auditory stimuli had a duration of 200 ms. The visual stimuli were two superimposed oppositely moving Gabor patches orientated such that motion was in the vertical-plane. The superposition, however, resulted in ambiguous motion—there was no overall physical upwards or downwards motion. The Gabor patches were spatially enveloped (sigma of spatial Gaussian equal to 1.07º) sinusoidal luminance gratings with a contrast of 0.05 per patch, a spatial frequency of 2.5 cycles per degree, and a temporal frequency of 6.25 Hz. In 90% of trials the ambiguous

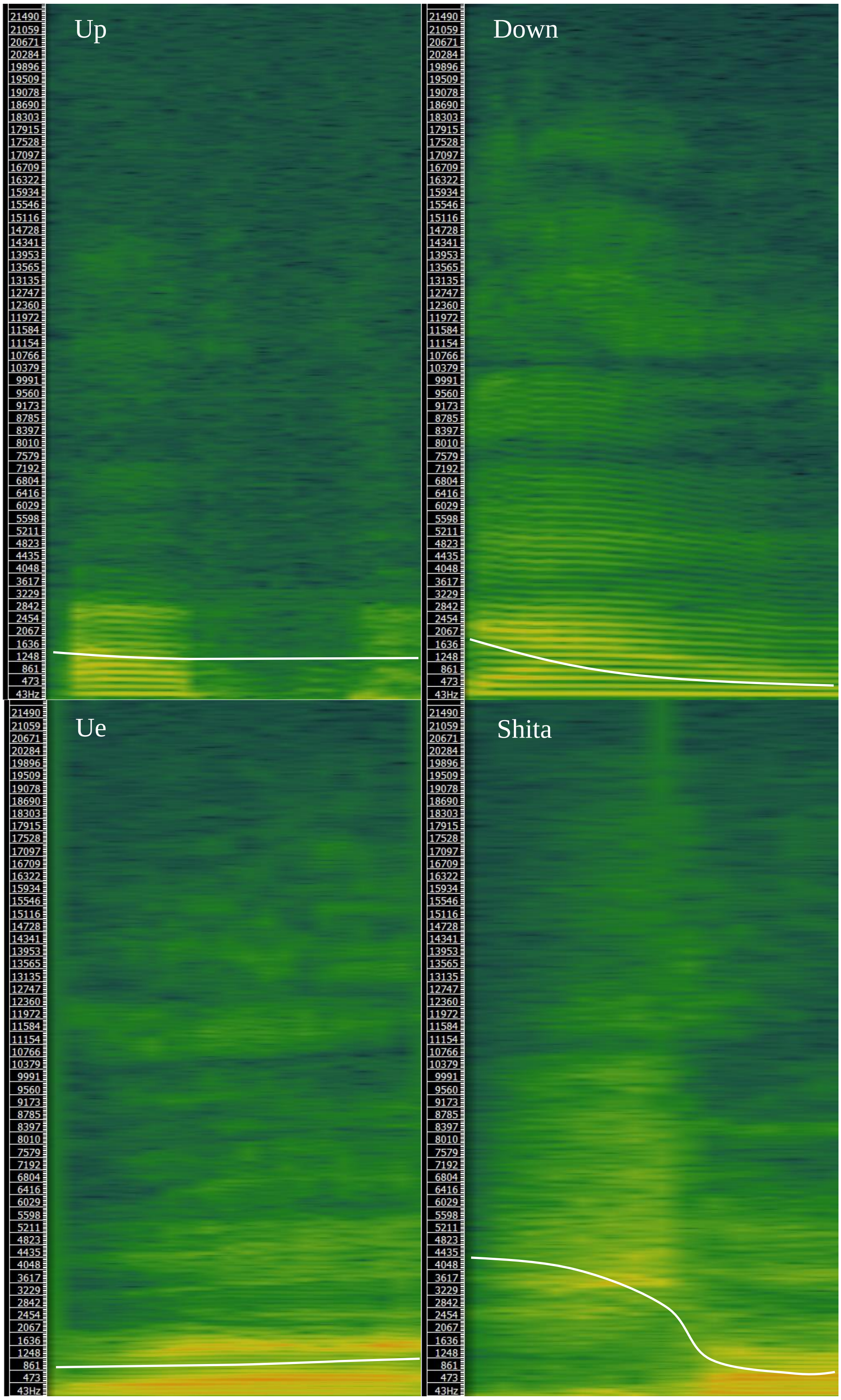

Figure 1: Spectrograms of all auditory stimuli used in Experiment 2, with line of best fit overlaid in white. Scale shown in black on the left-hand side (in Hz), with time on the x-axis and the full duration of the stimuli displayed (400 ms). Clockwise from top left: “up”, “down”, “ue”, and “shita” stimuli.

visual stimuli were used, with the remaining 10% consisting of a single upwards (5%) or downwards (5%) moving Gabor patch to keep the participants' attention. Visual and auditory stimuli were presented with five different SOAs: -300, -100, 0, +100, and +300 ms.

In Experiment 2, the visual stimuli were kept constant, but the auditory stimuli changed from pitches to speech. There were four auditory stimuli, consisting of the words "up" and "down" in English and Japanese ("ue" and "shita") each with a duration of 400 ms. Visual and auditory stimuli were presented simultaneously. Word stimuli were identical to those used in the study by Maeda et al. (2004), whilst pitch stimuli were recreated following the guidelines in the same paper. Spectral analysis of the word stimuli used in Experiment 2 can be viewed in Figure 1. Gabor patches and pitch stimuli were created and presented using Pyschtoolbox Version 3, a MATLAB toolbox (https://sccn.ucsd.edu/eeglab/index.php). The EEG script was created and presented using MATLAB (MATLAB and Statistics Toolbox Release 2015b, The MathWorks, Inc., Natick, Massachusetts, United States). Table 1 and Table 2 provide a summary of the stimuli used in these experiments and the details of each condition with stimuli timings respectively.

Overall, each condition had 100 trials resulting in a total of 1500 trials for Experiment 1 (three auditory conditions and five SOA conditions) and 400 trials for Experiment 2 (four auditory conditions). Throughout both experiments, the auditory and visual stimuli pairing was randomised. A fixation cross was presented in the centre of the screen for a duration of between 1400 to 1600 ms (randomised) prior to the auditory/visual condition. A question mark was presented at a randomised time point, appearing between 1000 – 1100 ms after the end of auditory/visual condition. Upon seeing the question mark, participants had 1500 ms to react by indicating their direction of motion perception, whereupon the fixation cross would appear immediately again.

Table 1

Summary of Experimental Stimuli

Stimuli	Type	Duration (ms)	Details	Experiment
Ascending	Auditory (Pitch)	200	Pure tone pitch glide (0.3–2.0 kHz)	1
Descending	Auditory (Pitch)	200	Pure tone pitch glide (2.0–0.3 kHz)	1
Broadband	Auditory (Pitch)	200	White noise constructed from randomisation of ascending pitch	1
“Up”	Auditory (Words)	400	Spoken English word	2
“Down”	Auditory (Words)	400	Spoken English word	2
“Ue”	Auditory (Words)	400	Spoken Japanese word (up)	2
“Shita”	Auditory (Words)	400	Spoken Japanese word (down)	2
Upwards Gabor	Visual	400	Single upwards moving Gabor	1+2
Downwards Gabor	Visual	400	Single downwards moving Gabor	1+2
Ambiguous Gabor	Visual	400	Two superimposed upwards/downwards moving Gabors	1+2

Table 2

Summary of Experimental Conditions

Condition	First Stimuli	Second Stimuli	Simultaneous Presentation?	Latency (ms)
SOA 1	Auditory (Pitch)	Visual	No	300
SOA 2	Auditory (Pitch)	Visual	No	100
SOA 3	N/A	N/A	Yes	0
SOA 4	Visual	Auditory (Pitch)	No	100
SOA 5	Visual	Auditory (Pitch)	No	300
English Words	N/A	400	Yes	0
Japanese Words	N/A	400	Yes	0

Design

The present study used a mixed design with both behavioural and EEG components. In the behavioural experiment, two within-subjects variables were manipulated: the sound condition accompanying the visual stimuli (ascending/descending pitch or words “up” or “down”), and the SOA—representing the time lag between visual and auditory stimulus presentation. The dependent variable was the participants judgement of whether the visual stimulus was upwards or downwards moving. In the EEG experiment, and additional dependent variable was the participants’ brain activity during the tasks.

Procedure

To test whether participants’ perception of visual motion was altered by the presence of auditory stimuli, their performance on a 2 AFC task asking them to determine the direction of motion of an ambiguous stimulus was recorded. Before completing the task, they were asked to fill out the consent form and three questionnaires (see Appendix B). Afterwards, a tape-measure was used to measure their head size so that an elastic head cap could be fitted. Next, the participants were prepared for EEG and sat in a dark, Faraday-caged room, approximately 80 cm away from a computer screen. A 20-inch LCD monitor running MATLAB was used to display the visual stimuli and 2 speakers placed on the ground in front of the participant delivered the auditory stimuli. Participants underwent five minutes of resting state EEG recording with eyes open, followed by five minutes with eyes closed. Then, instructions for the main tasks were presented and the participants were allowed 15 practice trials to become familiarised with the keyboard response and stimuli. The participants indicated via pressing the “1” or “4” key if they believed the visual stimuli was moving more downwards or upwards respectively. Participants were asked to respond as soon as they saw the prompt, which was a white question mark in the centre of the screen, and to remain as still as possible and to try not to blink during the blocks. Following the practice trials, Experiment 1 began, which consisted of 10 blocks with each block containing 150 trials. Each trial lasted 400 ms on average and each block lasted about 10 minutes. After each block, the participant was allowed a break where they could move or have a drink of water. Upon completing Experiment 1, a short break was administered before beginning the second experiment. Experiment 2 consisted of five blocks containing 80 trials each and lasted approximately 25 minutes. In total, the study lasted three hours on average, including 35 minutes to prepare the participant for EEG. They were exposed to 100 trials per condition and 1350 ambiguous visual stimuli trials in Experiment 1 and 360 ambiguous visual stimuli trials in Experiment 2 for a total of 1710 trials. The remaining trials consisted of unambiguous upwards and downwards moving visual stimuli which were included so that participants remained focussed.

EEG Recording and Analyses

EEG signals were recorded using 64 Ag-AgCl electrodes (Fp1, Fpz, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, O1, Oz, and O2) placed on the scalp according to the International 10/20 electrode placement system (Jasper, 1958). A BioSemi ActiveTwo amplifier was used to amplify the signal

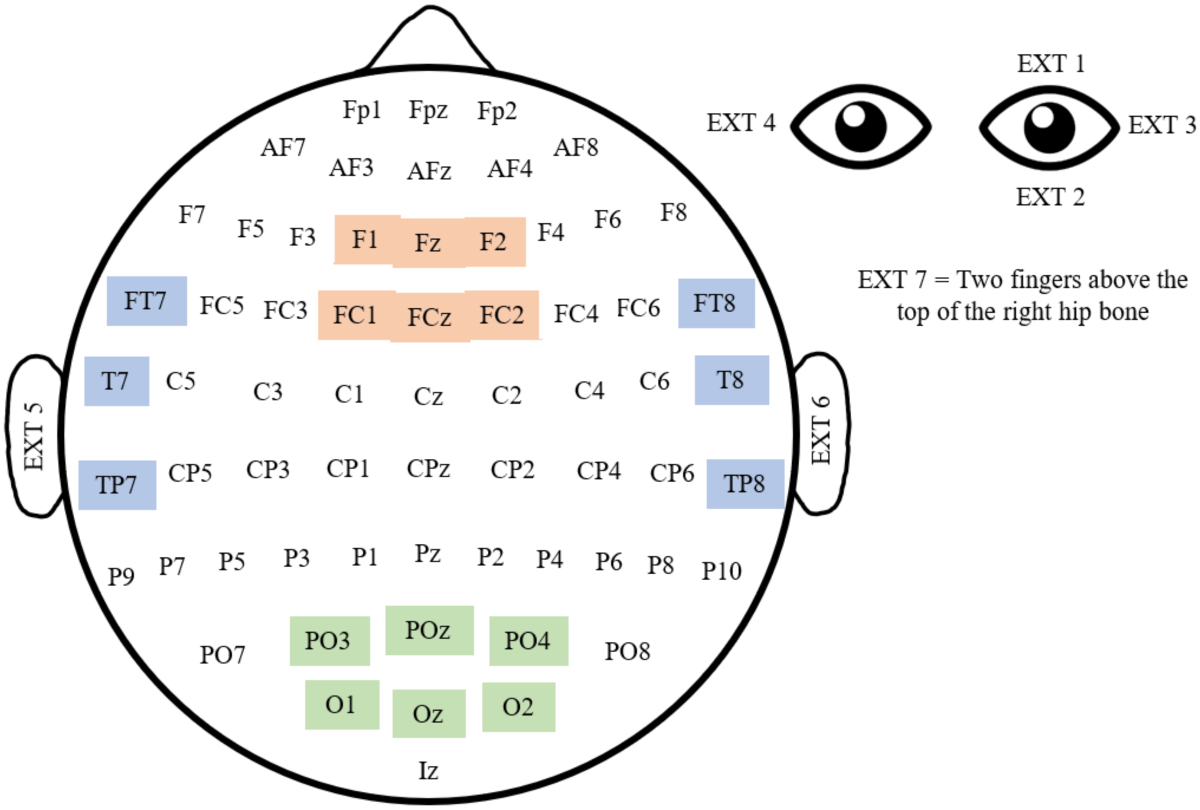

Figure 2: Map showing electrode placements. Frontal, Temporal, and Occipital electrode clusters used in ERP analysis are shown in orange, blue, and green respectively. External electrode placements are shown top right, including two reference electrodes placed on the earlobes. A total of 71 electrodes were used (64 + 7 externals).

(https://www.biosemi.com/products.htm). A computer running Microsoft Windows was used in conjunction with BioSemi ActiView v7.03 (https://www.biosemi.com/download.htm) to record the EEG data as a “.bdf” file and to record triggers. In total, seven external electrodes were used. The vertical electro-oculogram (VEOG), which was used to detect eye-blinks, was recorded from electrodes placed below and above the left eye, whilst two electrodes were used to record the horizontal electro-oculogram (HEOG), which tracked saccades. Two reference electrodes were placed on the left and right earlobe, and one electrode to measure heart rate was placed two fingers above the top of the right hip bone. Scalp electrodes were clustered bilaterally into three regions for analysis—*frontal*, *temporal*, and *occipital*—each containing an average of 6 electrodes. A map of all electrode placements and cluster regions can be seen in Figure 2. Impedances were kept below 5 kΩ. The EEG recording used a sampling frequency of 1024 Hz, which was later re-sampled to 512 Hz.

Pre-processing. Pre-processing was done using EEGLab, a MATLAB toolbox (Delorme & Makeig, 2004) and ERPLab (Lopez-Calderon & Luck, 2014) was used for further data analysis. The scalp electrodes were first re-referenced to the mean of the two earlobe electrodes and a 0.5 Hz high pass filter and 48-52 Hz notch filter were applied to the continuous data. Independent Component Analysis (ICA) was conducted on the data and any bad channels were then interpolated. Artefact rejection was carried out manually to correct for eye-blink and saccade related artefacts. For each of the five SOAs, the data was epoched from -1500 to +1500 ms, time-locked to the onset of the latter of the two stimuli presented. For ERP analysis, a baseline correction was applied, anchored to a 500 ms period before stimuli presentation and during which the fixation cross was being displayed. An additional 35 Hz low pass filter was also used. Trials with artefacts exceeding amplitudes of ± 100 µV on any channel were rejected and an artefact rejection cut-off of 30% across each participant was used to determine whether to omit participant data. On average, 13.1% of all trials were rejected and six participants were omitted. The trials were averaged over each condition for each participant, and this average was used for the later statistical analysis. Finally, a grand average across participants was calculated and used in the plotting of ERPs and topographic scalp maps.

Results

Two sets of analyses were carried out on the raw data collected from Experiment 1 and 2. Firstly, a behavioural analysis of the participants' responses to the task, to discover if the illusion was truly present and if so, which sets of parameters affected it the most. Secondly, pre-processing, followed by event related potential (ERP) analysis was conducted on the EEG

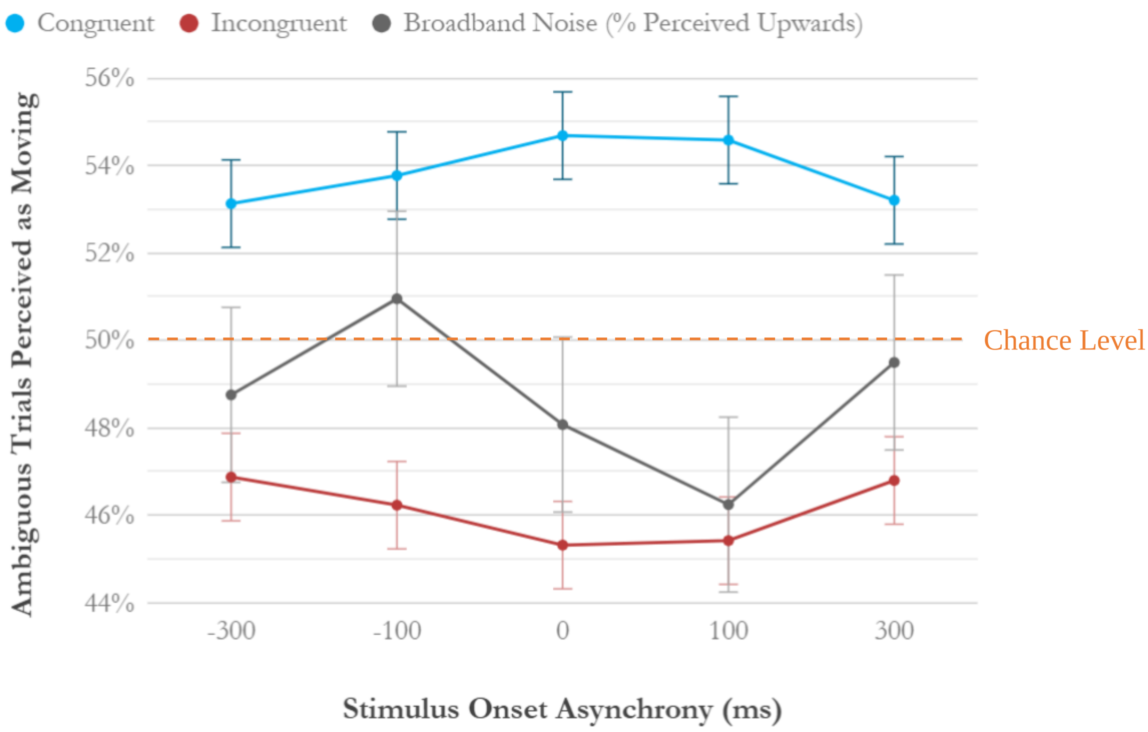

Figure 3: Experiment 1 behavioural results. Average percentage of trials perceived as moving in congruent (blue) or incongruent (red) directions for each of the five SOA conditions (-300, -100, 0, +100, +300 ms). Percentages are values averaged over all trials and all participants. For the auditory stimuli condition "Broadband Noise", percentage of trials perceived as moving upwards is displayed. Note that negative SOAs represent auditory stimuli being presented *before* visual stimuli and vice versa for positive SOAs. Congruent trials are those in which the participant perceived motion, for an ambiguous visual stimuli, in the same direction (up or down) as the auditory stimuli (ascending or descending). Error bars are plotted as one standard error of the mean.

data to look for patterns of activity during certain time points in each trial. The details of these two sets of analyses are presented in the subsequent sections.

Behavioural Results

After pre-processing the raw EEG data, and completing artefact rejection, six participants were identified as having data that was too noisy to analyse and two participants possessed incomplete EEG data recording. Therefore, a total of eight participants were omitted and the final sample analysed consisted of 23 individuals aged between 20 and 36 years old (M = 25.9, SD = 4.3; 15 females). The average percentage of trials where the participant did not reply in time or pressed an incorrect button was calculated for Experiment 1 (missed = 2.9%; incorrect = 2.6%) and Experiment 2 (missed = 1.6%; incorrect = 2.6%).

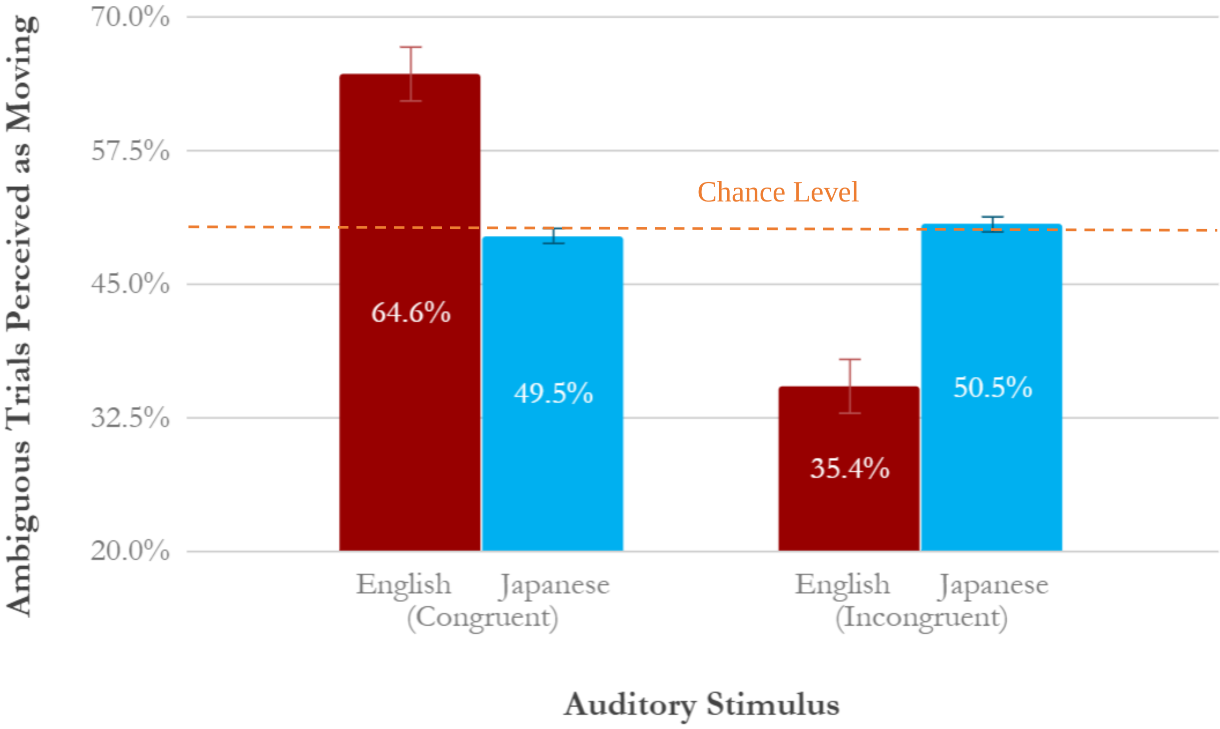

Figure 4: Experiment 2 behavioural results. Average percentage of trials perceived as moving in congruent and incongruent directions for English (red) and Japanese (blue) words. Percentages are values averaged over all trials and all participants. Auditory stimuli consist of the spoken words "up" and "down" (English) and "ue" and "shita (Japanese) and last approximately 400 ms. Congruent trials are those in which the participant perceived motion for an ambiguous visual stimuli in the same direction (up or down) as the auditory stimuli (semantic meaning of up or down). Error bars are plotted as one standard error of the mean. Chance level for motion perception is 50%.

Participants were separated by condition and the average percentages of congruent and incongruent decisions for the corresponding conditions were calculated (see additional materials provided alongside this paper for data tables). There were five different SOAs (-300, -100, 0, +100, +300 ms), and three different auditory conditions (ascending, descending, and broadband) resulting in a total of 15 conditions for Experiment 1; for Experiment 2, there was one SOA and four different auditory conditions ("up", "down", "ue", and "shita") resulting in a total of 4 conditions. The results of the average percentages of congruent and incongruent decisions for all of these conditions is plotted in Figure 3 and Figure 4 for Experiment 1 and Experiment 2 respectively.

Figure 3 shows the average percentage of ambiguous trials perceived as moving in a congruent (blue line) or incongruent (red line) direction, where congruent represents perceived motion in the same direction as the auditory stimuli. The percent of broadband noise stimuli eliciting motion perception in the upwards direction is also shown in grey and values reside close to the chance level (50%) across all SOA conditions. Error bars displayed represent one standard error of the mean and are calculated on an individual basis for each SOA and congruency. Upon visual inspection, a slight difference between average percentage of congruent and incongruent trials is present, with congruent trials occurring 8% more of the time than incongruent trials.

To determine if this difference was significant, a paired-samples t-test was conducted comparing percentage of trials perceived in a congruent and incongruent direction against themselves and against the chance value of 50%, using a Bonferroni correction and $\alpha = 0.01$. There was a significant difference in the percentages for congruent and incongruent decisions across all SOAs, and effect sizes ranged from 0.47 – 0.63. The results of these t-tests and calculated effect sizes are displayed in Table 3. In Table 3, mean values and standard deviations for congruent (M_{con}, SD_{con}) and incongruent trials (M_{incon}, SD_{incon}) are displayed alongside t values and their associated p-values, as well as a post-hoc estimate of effect size. The largest effect of congruency occurs with an effect size of 0.63 at SOA 3, where stimuli are presented simultaneously; $t(31) = 4.40$, $p < .001$. There was also a significant difference, across all SOAs, in the percentages for congruent and incongruent decisions as compared to the chance level (50%), whereas no significant difference was found when comparing broadband noise trial decisions against chance. Taken together, findings from Figure 3 and the statistics in Table 3 can be summarised by a main effect of congruency, strongest at zero latency between auditory and visual stimuli, mediated by ascending or descending pitches and not broadband noises.

Table 3

Statistical t-tests for Experiments 1 & 2

Comparison	M_{con}	SD_{con}	M_{incon}	SD_{incon}	t	df	p	Effect Size
Congruent vs Incongruent (SOA 1)	.532	.062	.468	.062	2.85	30	.008	.468
Congruent vs Incongruent (SOA 2)	.540	.054	.461	.054	3.98	30	<.001	.595
Congruent vs Incongruent (SOA 3)	.549	.061	.451	.061	4.40	30	<.001	.632
Congruent vs Incongruent (SOA 4)	.548	.063	.452	.063	4.15	30	<.001	.610
Congruent vs Incongruent (SOA 5)	.533	.053	.467	.053	3.44	30	.002	.539
Congruent vs Incongruent (English)	.646	.129	.354	.129	6.65	30	<.001	.753
Congruent vs Incongruent (Japanese)	.495	.040	.505	.040	-0.67	30	.506	.122

Similar analyses were conducted for Experiment 2, which investigated whether hearing the words “up” and “down” in English and Japanese could influence perception of the direction of motion of the ambiguous visual stimuli. Figure 4 displays the average percentage of congruent and incongruent trials for English (red bars) and Japanese (blue bars) auditory conditions, with error bars at the level of one standard error of the mean. Again, paired-samples t-tests were conducted between the percentages of congruent and incongruent trials and is displayed in Table 3. Visual inspection suggests a much larger effect of congruency is present for the word conditions as compared to the pitch conditions, which is only present for the English words condition—64.6% congruency compared to 35.4% incongruency for a chance level of 50%. Indeed, upon performing paired-samples t-tests between congruent conditions for the English stimuli, we observe a significant effect of congruency which is stronger than for the pitch stimuli in Experiment 1; $t(30) = 6.65$, $p <$ .001. Japanese word conditions, however, provide no significant congruency effect; $t(30) =$ -

0.67, p = .506, and so congruent and incongruent decisions are not significantly different to each other or to chance. When only individuals whose native language was English are considered (N = 13), the percentage of congruent trials increases from 64.6% to 69.2%. Similarly, when only males (N = 10) are considered, the percentage of congruent trials becomes 68.1%, compared to 63.4% for only females (N = 21). However, there is no considerable change when considering age group. Taken together, the behavioural results from Experiment 2 find that there is a larger significant effect of congruency for words compared to pitches, which vanishes for Japanese words.

EEG Results

Following the behavioural analysis, ERP analysis was conducted on the pre-processed (see Pre-processing section) EEG data. This was comprised of three main stages—plotting the grand average waveform ERP responses over time, creating topographic scalp maps for each individual ERP component, and applying statistical techniques to test the ERPs for significance. These three stages will be discussed in order over the subsequent paragraphs.

Waveform ERP Response. Firstly, to identify ERP components present in the data, the amplitude (in µV) of the recorded signal was plotted over all epoched conditions. A time window of 1000 ms pre- and post-stimuli (visual) was chosen for display purposes, and error bars at one standard error of the mean were overlaid. Two conditions, congruent and incongruent, were plotted for three different regions of the brain—*frontal*, *temporal*, and *occipital*—each consisting of an average of 6 electrodes. For Experiment 1, this culminated in 15 plots, showing five different SOAs over three brain regions, which are displayed as sub-figures in Figure 5. Overlaid, are the onsets of the visual and auditory stimuli in green and blue respectively. Throughout all SOA conditions, the figures remain anchored to the visual stimuli, which always appears at the origin, whilst the auditory stimuli move between latencies of 300 and 100 ms either side of the origin. Baseline corrections were applied using a 500 ms baseline between -1400 to -900 ms when the fixation cross was on-screen. Known visual and auditory ERP components were used to label the possible ERP components present in these figures, however, due to the crossmodal aspect of this experimental design, it is likely that the ERPs found are a superposition of both auditory and visual components.

Nevertheless, it appears that several ERP components are present for Experiment 1, including C1, N1, P2, N2, P3, and possibly N4 and P6.

The same process was completed for Experiment 2, producing six figures in total—two conditions (English or Japanese words) across the same three separate brain regions, as displayed in Figure 6. Because the stimuli were presented simultaneously, auditory and visual stimuli markers always appear at the origin. Again, baseline corrections were applied using a 500 ms baseline between -1400 to -900 ms when the fixation cross was on-screen. There appear to be several ERP components present for Experiment 2, including C1, N1, P2, N2, P3, and possibly P6. The components observed in Experiments 1 and 2 will be discussed in more detail during the Discussion section.

Topographic Scalp Maps. Once the known ERP components had been identified using the waveform ERP response figures, the characteristic time windows for those components (Sur & Sinha, 2009; Woodman, 2010) were used to plot topographic scalp maps. A topographic scalp map was plotted for each individual ERP component, as displayed in the subfigures within Figure 7 and Figure 8 for Experiment 1 and Experiment 2 respectively. For each condition, three topographic figures were created—congruent, incongruent, and difference (congruent minus incongruent) responses—using the same colour bar scale ranging from -3 to +3 µV to assist direct comparisons between figures. The individual ERP components are shown at the top of each topographic figure and the arrow of time and the time-window for each ERP component are shown at the bottom, including the time elapsed since the first *and* second stimuli in the case of non-zero latency SOAs.

Statistical Analysis. To determine whether differences between the ERP responses for congruent and incongruent trials was statistically significant, a two by three repeated measures ANOVA was conducted over the entire dataset for each ERP identified. The first factor was the condition congruency, with two levels (congruent or incongruent) and the second factor was region, with three levels (frontal, temporal, and occipital clusters). This ANOVA was conducted for the individual ERP components identified in each SOA and word condition, resulting in 38 separate ANOVAs. The reason for this high number of statistical tests was the fact that separate SOAs may contain similar ERP components to each other but as they are transposed in time, the comparison time-window is not fixed for a chosen ERP component, and different stimuli offsets brings the additional effect of unique summations of auditory and visual ERPs. A Bonferroni correction was applied, and the mean scores for congruency was compared across region, and main effects of congruency, region, and the congruency-region interaction were calculated. The results of this statistical analysis are presented in the following sections, separated by condition, and interpreted in the Discussion section. Table 4 (see Appendix C) contains the entire ERP statistical analysis for Experiments 1 and 2 and is referred to throughout.

SOA 1. Seven ERP components (C1, N1, P2, N2, N1, P2, and P3) were tested for congruency, region, and interaction effects. No effects of congruency were found, with the closest effects at the N1 and P2 components at the time periods 130 – 190 ms and 170 – 230 ms respectively; $F(1,22) = 2.16$, $p = .156$, $\eta^2 = .089$ and $F(1,22) = 2.02$, $p = .169$, $\eta^2 = .084$. All seven components displayed effects of brain region at the $p < .001$ significance level. No interaction effects were discovered, however, the aforementioned N1 and P2 components displayed the highest non-significant effect.

SOA 2. Five ERP components (C1, P2, N2, P3, and N4) were tested. No effects of congruency or interactions were discovered, however, the P2 component had a near significant interaction effect; $F(2,44) = 3.02$, $p = .059$, $\eta^2 = .121$. All components were found to have significant effects of brain region at the $p < .01$ level.

SOA 3. Five ERP components (N1, P2, N2, P3, and P6) were tested. No effects of congruency or interactions were discovered. All components except N1 were found to have effects of brain region at varying levels of significance.

SOA 4. Four ERP components (C1, P2, N2, and P3) were tested. No effects of congruency or interactions were discovered, however, the N2 component had a near significant interaction effect; $F(2,44) = 2.87$, $p = .068$, $\eta^2 = .115$. Significant effects of brain region at the $p < .01$ level were found for all four components.

SOA 5. Five ERP components (N1, P2, N4, N2, and P3) were tested. No effects of congruency or interactions were discovered, however, the N4 component had a near significant interaction effect; $F(2,44) = 2.91$, $p = .065$, $\eta^2 = .117$. Similar to SOA 3, all components except N1 were found to have significant effects of brain region, this time at the $p < .05$ level.

English Words. Six ERP components (C1, N1, P2, N2, P3, and P6) were tested. One component (P3) displayed an effect of congruency and another (P6) showed a near-effect; $F(1,22) = 5.88$, $p < .05$, $\eta^2 = .211$ and $F(1,22) = 3.20$, $p = .087$, $\eta^2 = .127$ respectively. One interaction effect was present for the P6 component; $F(2,44) = 2.71$, $p < .001$, $\eta^2 = .110$, and on further inspection, the difference was between the occipital and frontal brain regions. All components except N2 were found to have significant effects of brain region at the $p < .01$ level.

Japanese Words. The same six ERP components (C1, N1, P2, N2, P3, and P6) were tested. All components except P6 displayed effects of congruency at varying degrees of significance from $p = .05$ to .001. All except the N2 component displayed effects of brain region. There were no significant interaction effects present for this condition.

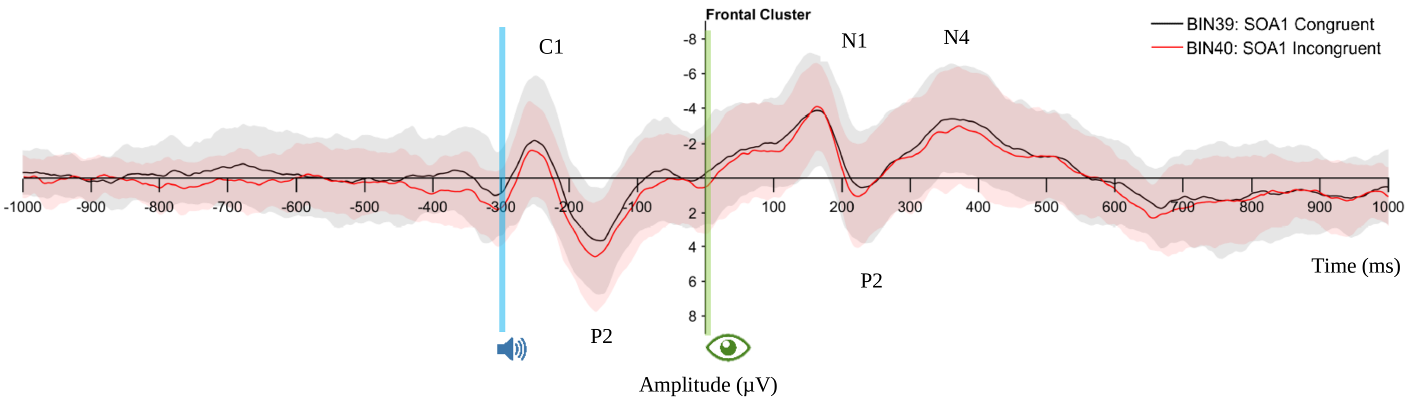

Figure 5a: SOA 1 condition; frontal cluster.

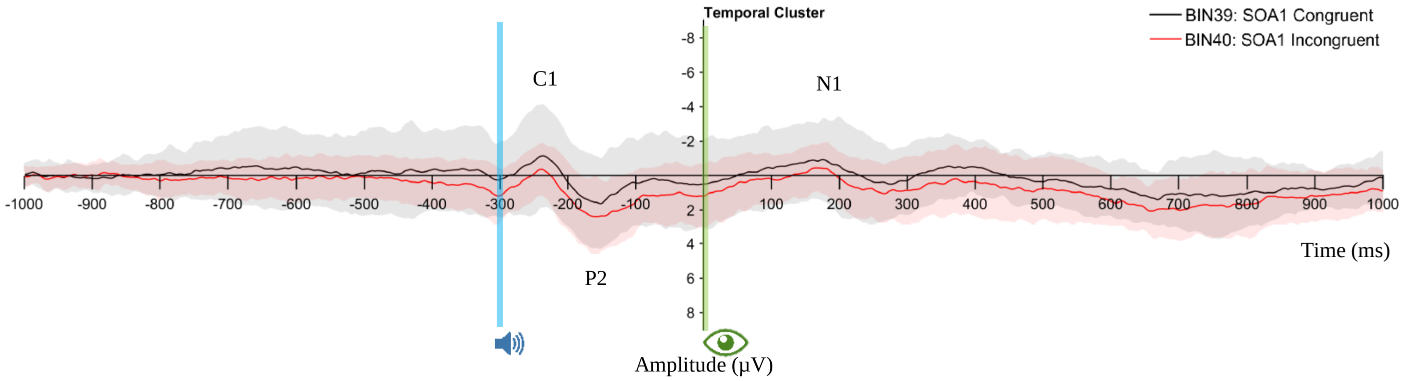

Figure 5b: SOA 1 condition; temporal cluster.

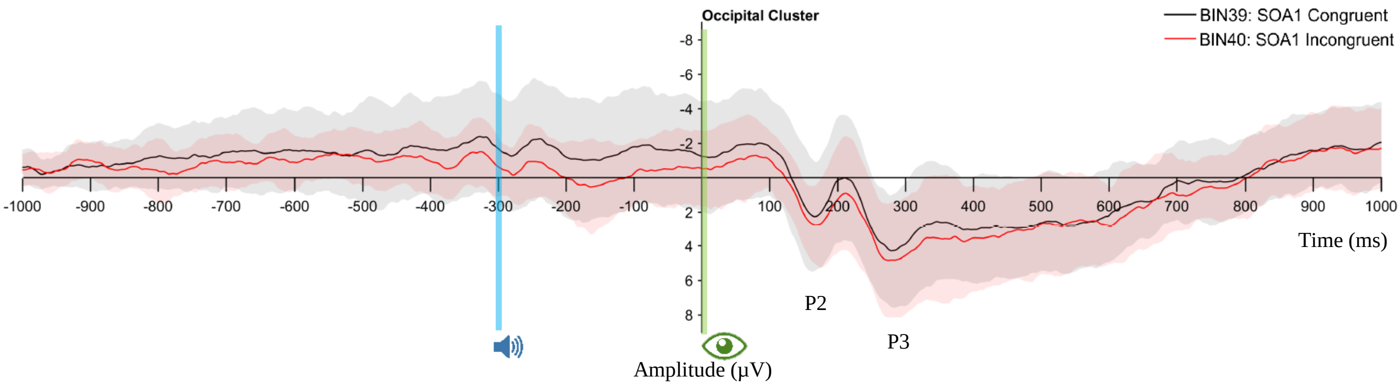

Figure 5c: SOA 1 condition; occipital cluster.

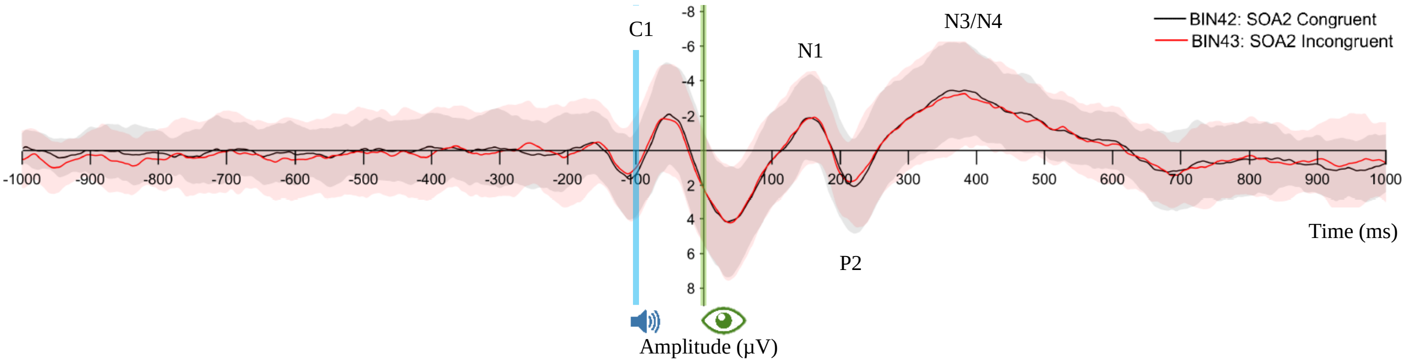

Figure 5d: SOA 2 condition; frontal cluster.

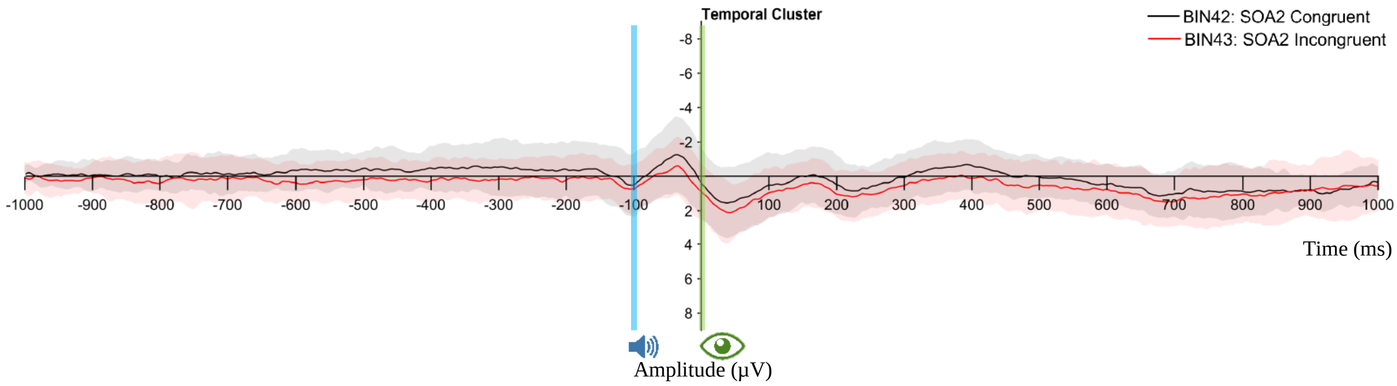

Figure 5e: SOA 2 condition; temporal cluster.

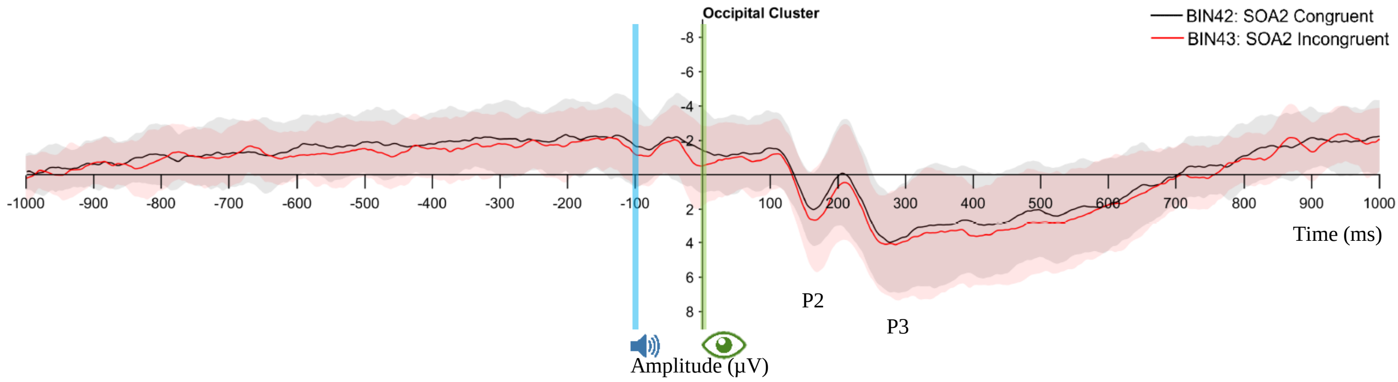

Figure 5f: SOA 2 condition; occipital cluster.

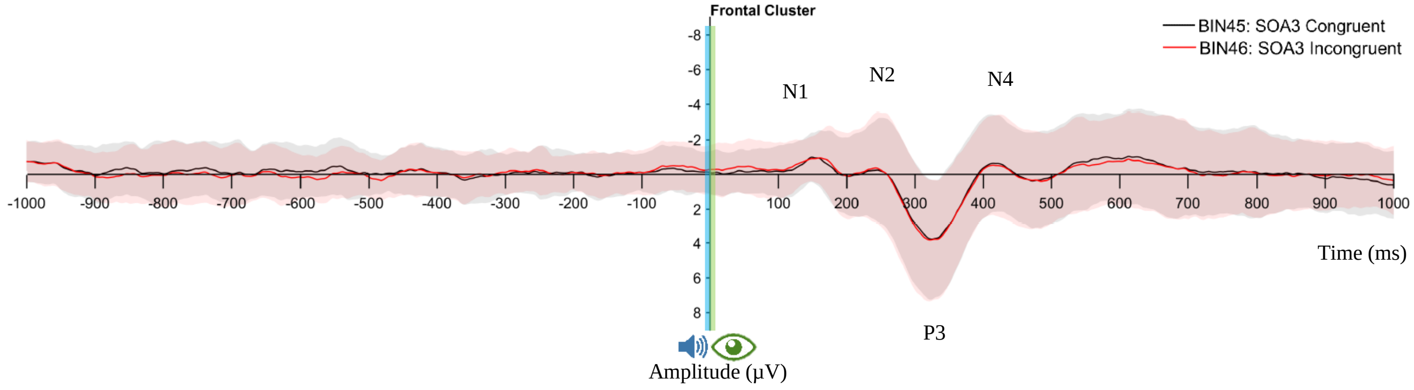

Figure 5g: SOA 3 condition; frontal cluster.

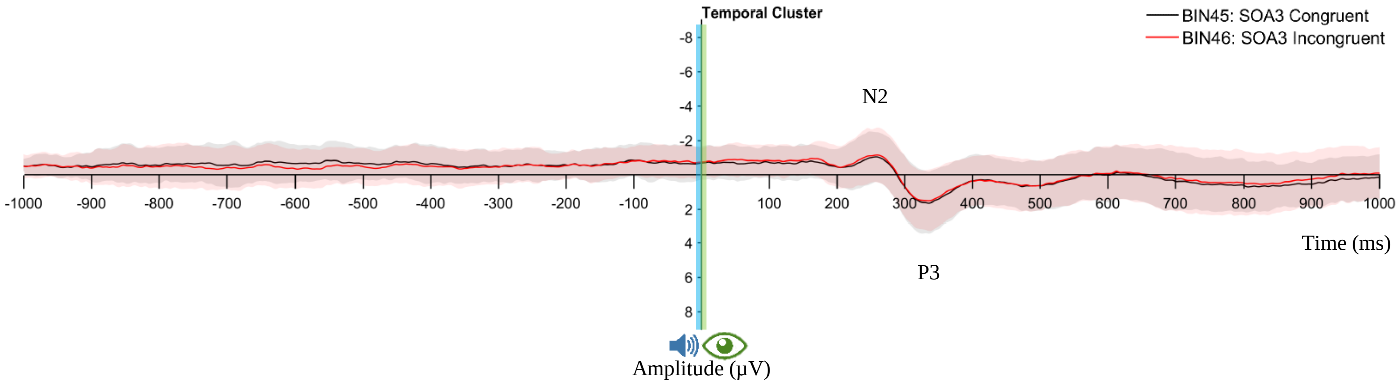

Figure 5h: SOA 3 condition; temporal cluster.

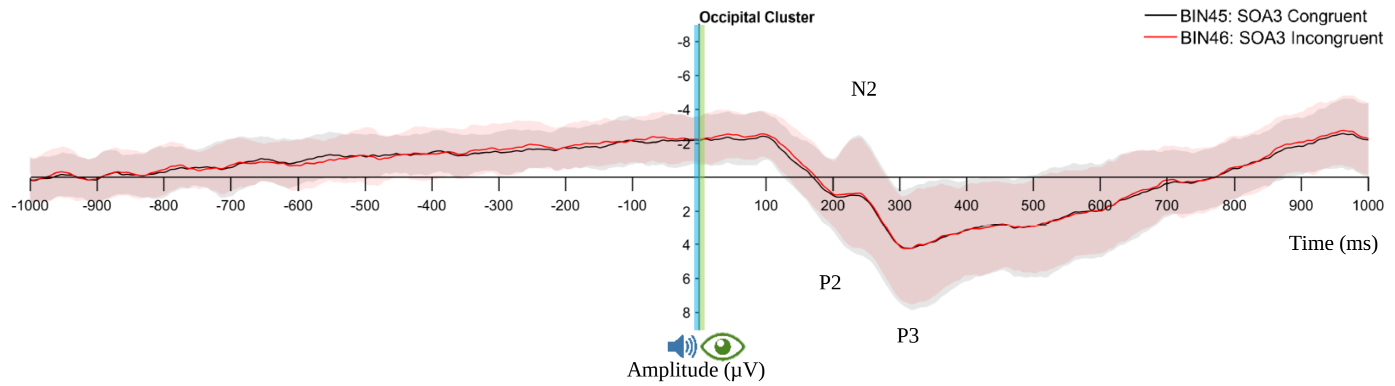

Figure 5i: SOA 3 condition; occipital cluster.

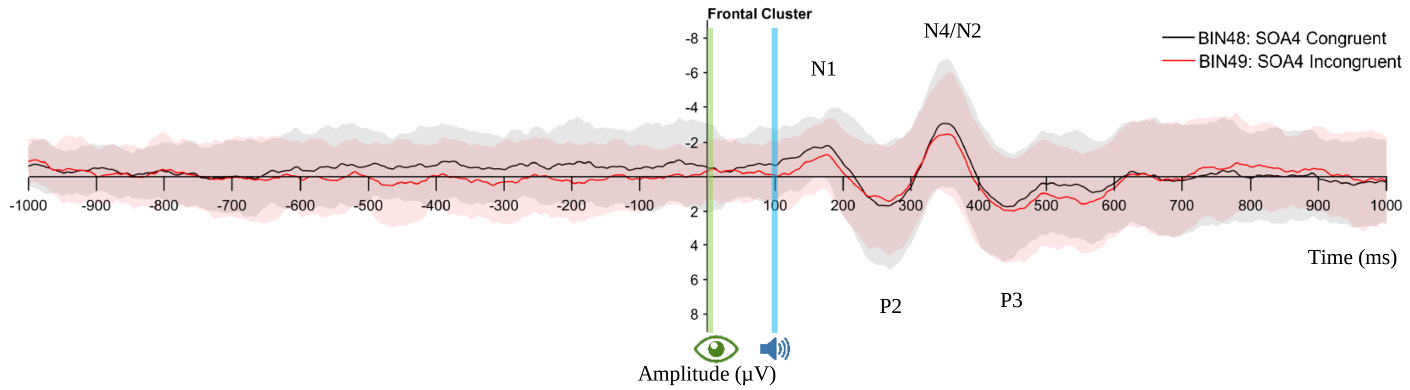

Figure 5j: SOA 4 condition; frontal cluster.

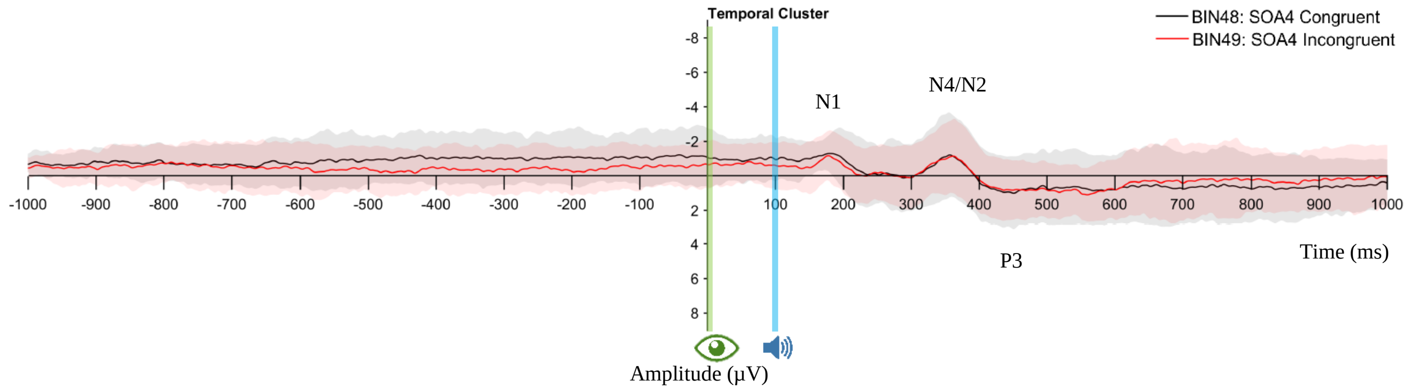

Figure 5k: SOA 4 condition; temporal cluster.

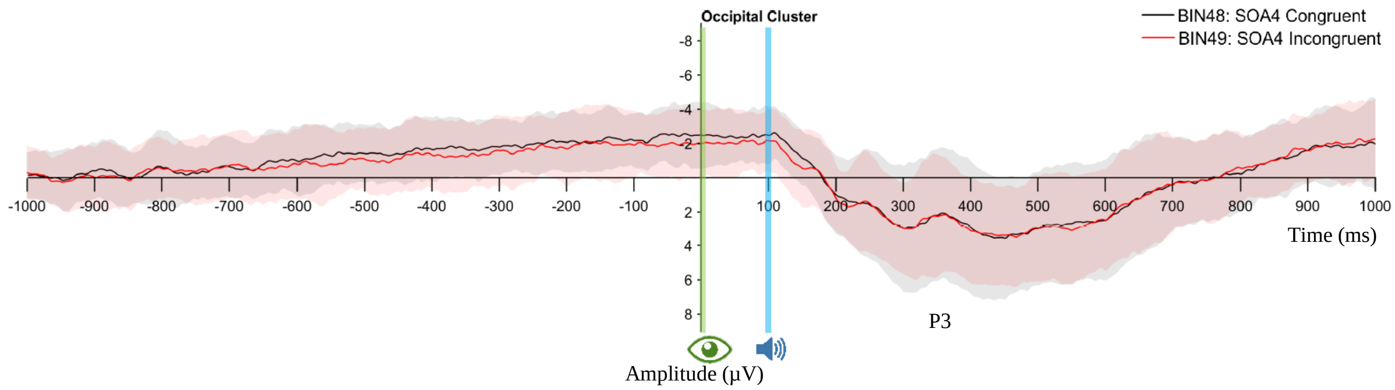

Figure 5l: SOA 4 condition; occipital cluster.

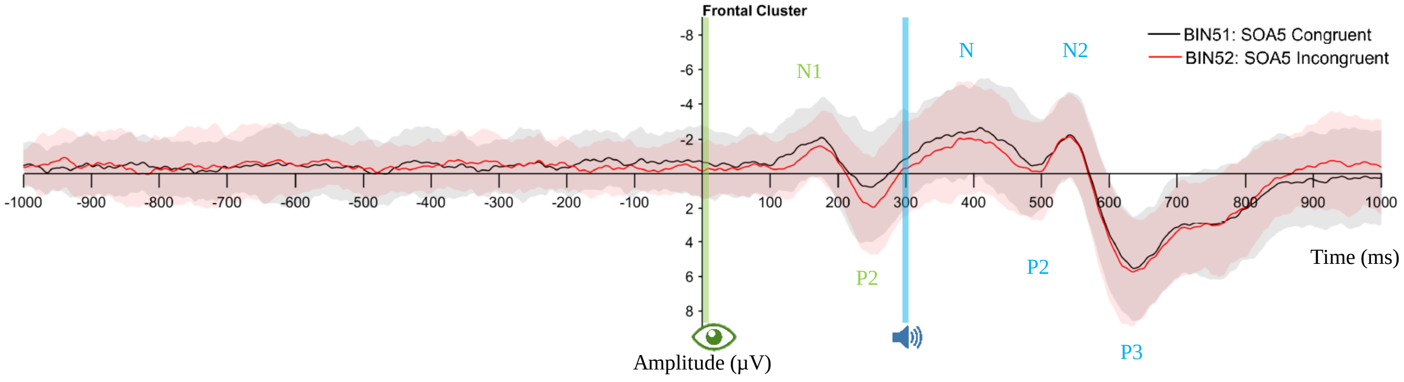

Figure 5m: SOA 5 condition; frontal cluster.

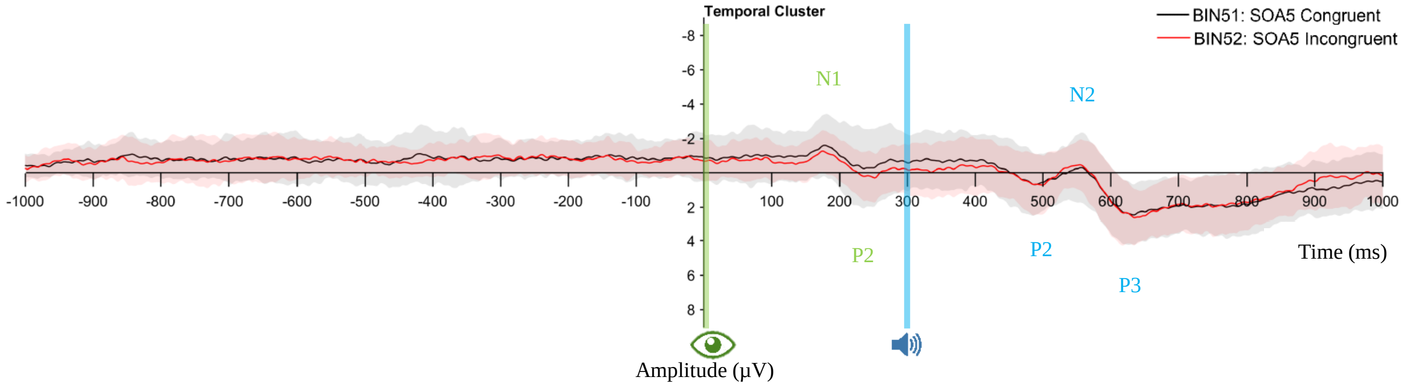

Figure 5n: SOA 5 condition; temporal cluster.

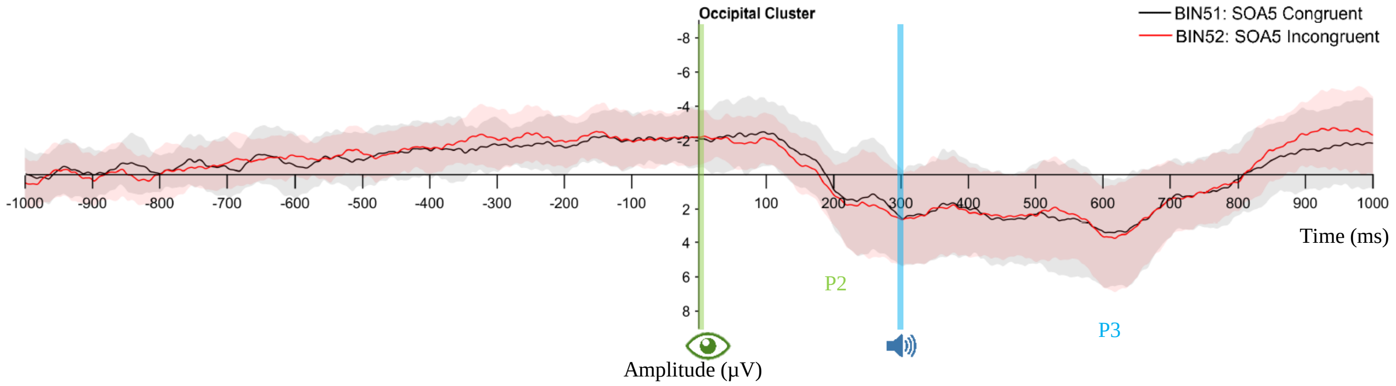

Figure 5o: SOA 5 condition; occipital cluster.

Figure 5: Experiment 1 EEG results. Grand average ERP responses for congruent (black line) and incongruent (red line) responses to ascending and descending pitch conditions for each of the *five SOA conditions* (-300, -100, 0, +100, +300 ms), shown 1000 ms pre- and post- stimulus presentation (auditory/visual) for the frontal, temporal, and occipital clusters (see Method section). Auditory (blue) and visual (green) stimuli onset are shown as vertical lines and a small icon. Possible ERP components are overlaid (see Figure 7 for topographic plots), however, a mixture of auditory and visual components may be present. Uncertainties at the one standard error of the mean level are plotted in the appropriate colour for congruent/incongruent trials. ERPs had completed artefact rejection (30% cut-off), notch filtered (48–52 Hz), and baseline corrected using a 500 ms period pre-stimulus during the presentation of the fixation cross (-1400 to -900 ms).

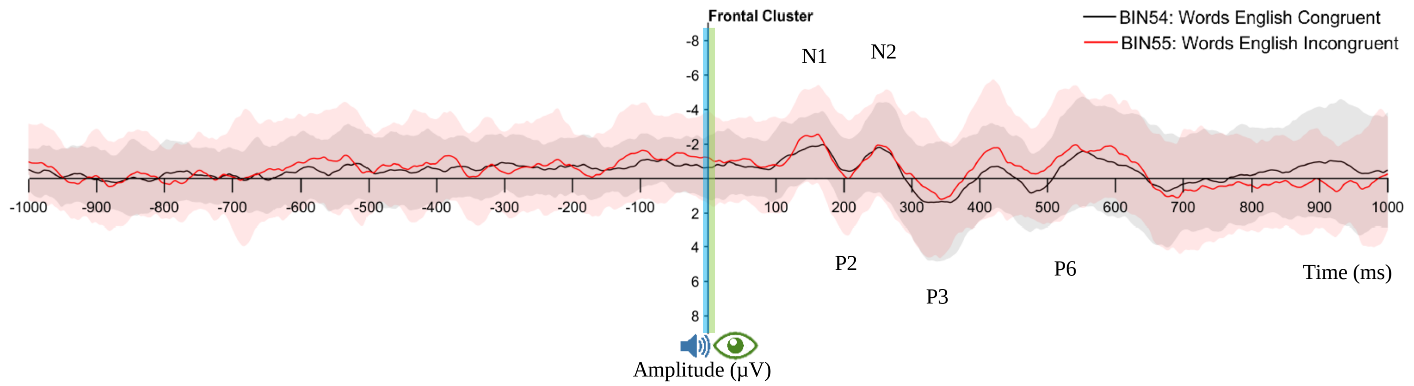

Figure 6a: English auditory condition; frontal cluster.

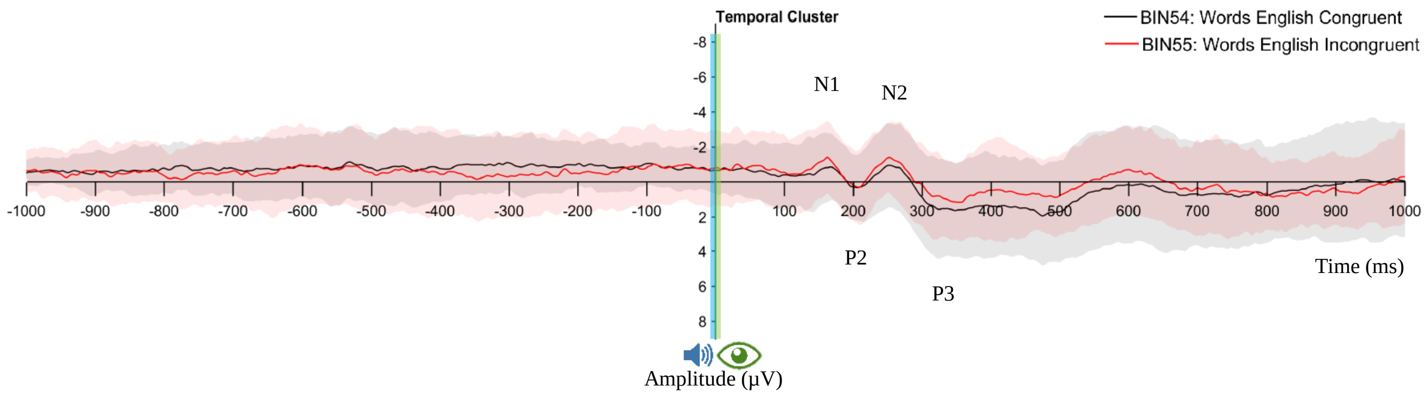

Figure 6b: English auditory condition; temporal cluster.

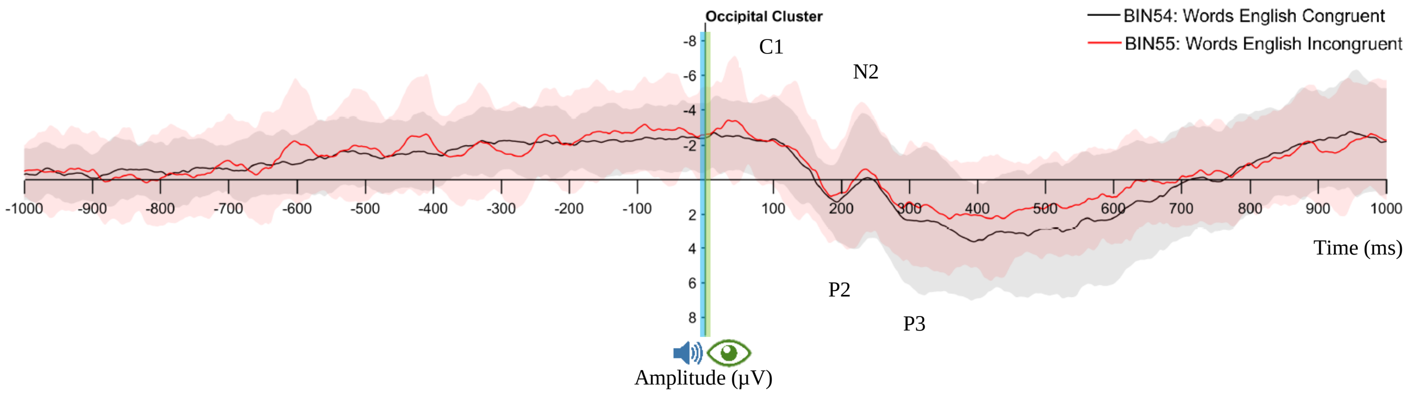

Figure 6c: English auditory condition; occipital cluster.

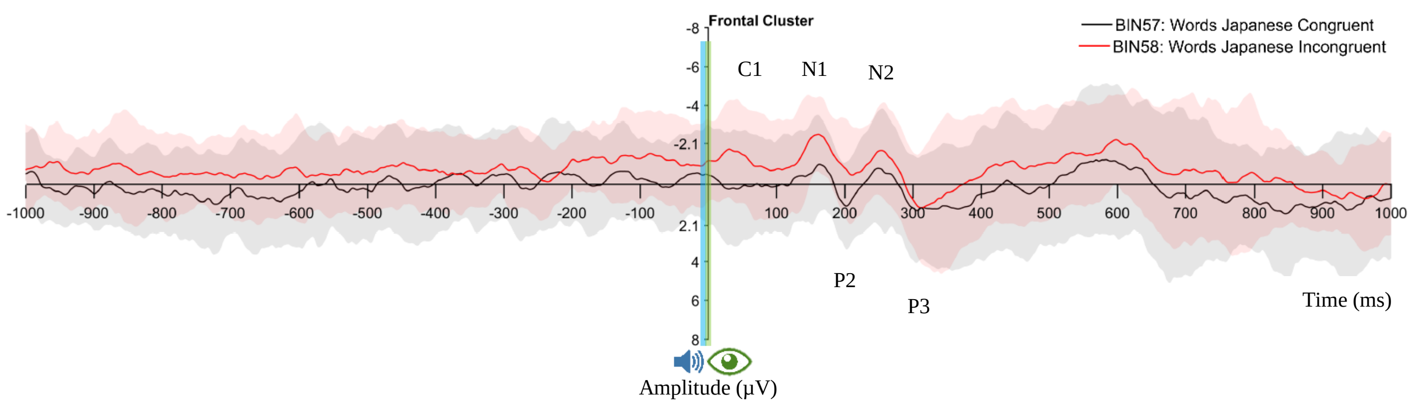

Figure 6d: Japanese auditory condition; frontal cluster.

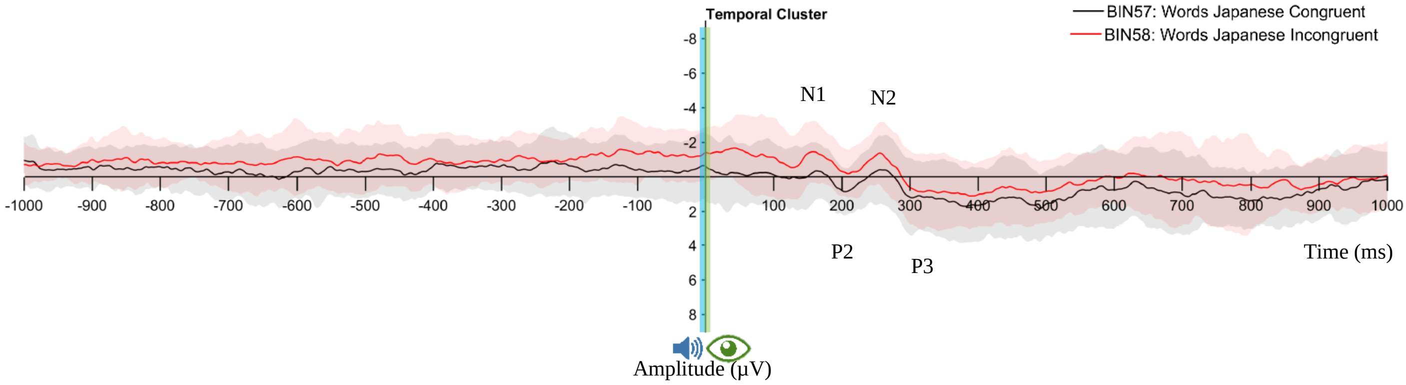

Figure 6e: Japanese auditory condition; temporal cluster.

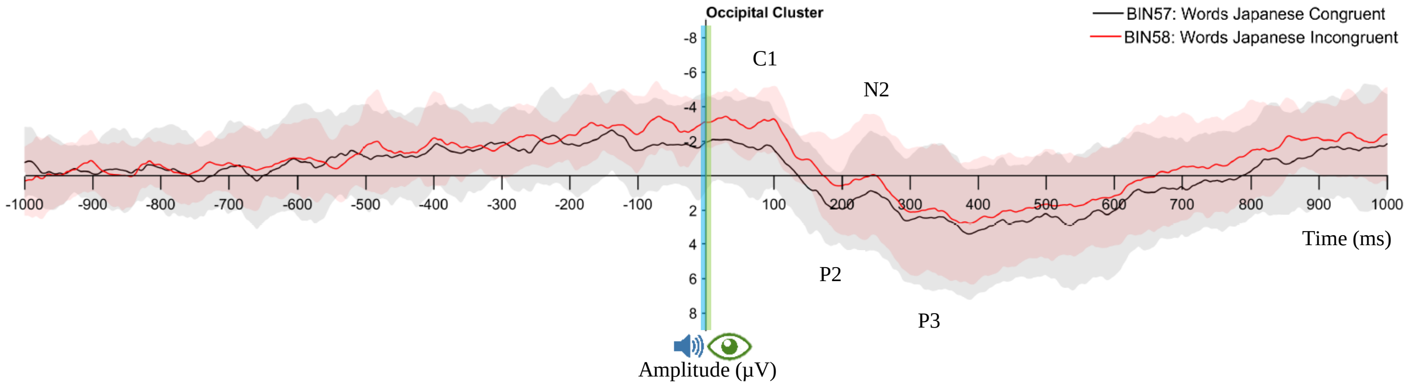

Figure 6f: Japanese auditory condition; occipital cluster.

Figure 6: Experiment 2 EEG results. Grand average ERP responses for congruent (black line) and incongruent (red line) responses to *English* and *Japanese conditions*, shown 1000 ms pre- and post-stimulus presentation (auditory and visual) for the frontal, temporal, and occipital clusters (see Method section). Auditory (blue) and visual (green) stimuli onset are shown as vertical lines and a small icon. Possible ERP components are overlaid (see Figure 8 for topographic plots), however, a mixture of auditory and visual components may be present. Uncertainties at the one standard error of the mean level are plotted in the appropriate colour for congruent/incongruent trials. ERPs had completed artefact rejection (30% cut-off), notch filtered (48–52 Hz), and baseline corrected using a 500 ms period pre-stimulus during the presentation of the fixation cross (-1400 to -900 ms).

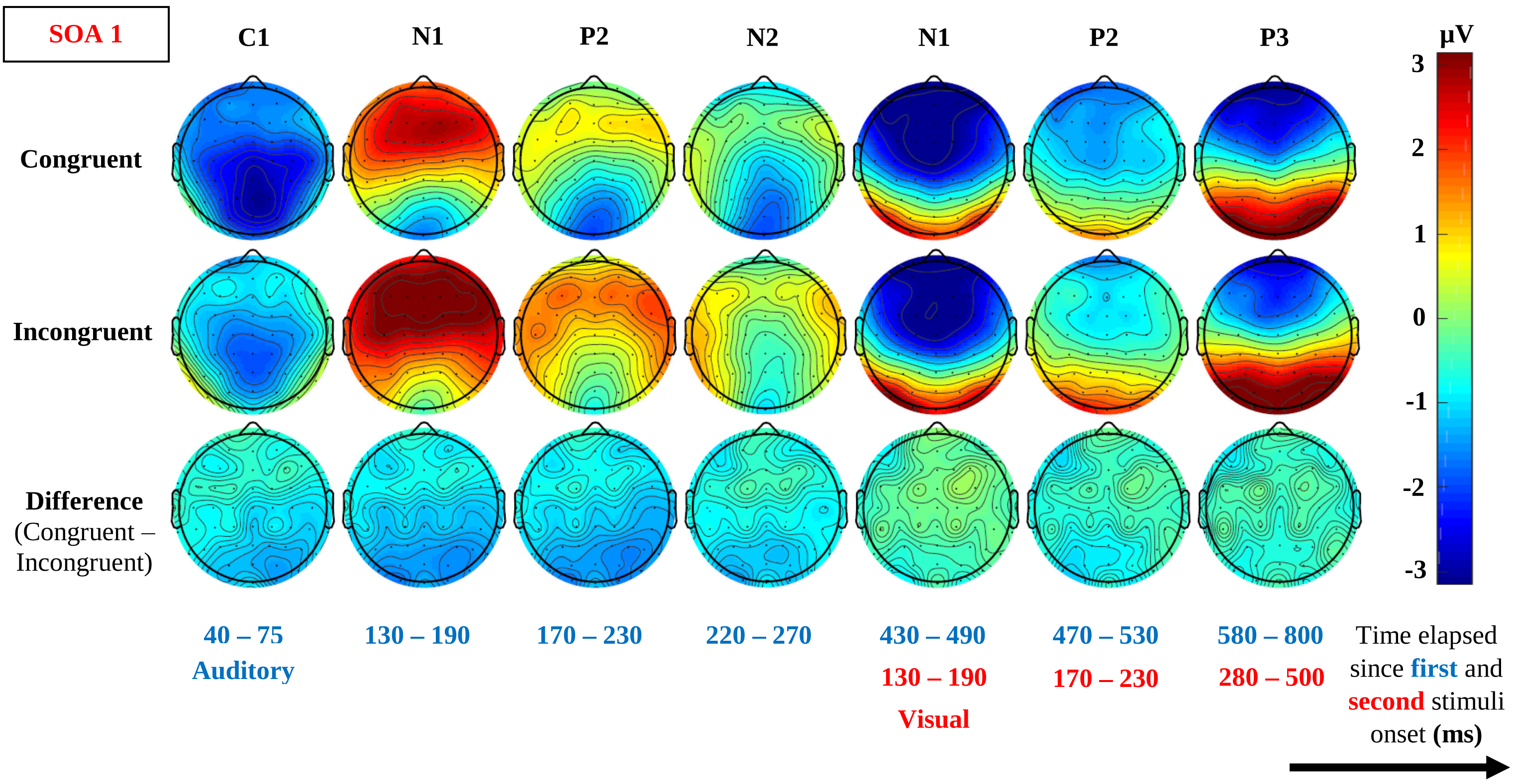

Figure 7a: Experiment 1; Grand average topographic maps for *SOA 1* with congruent, incongruent, and difference (congruent minus incongruent) conditions displayed.

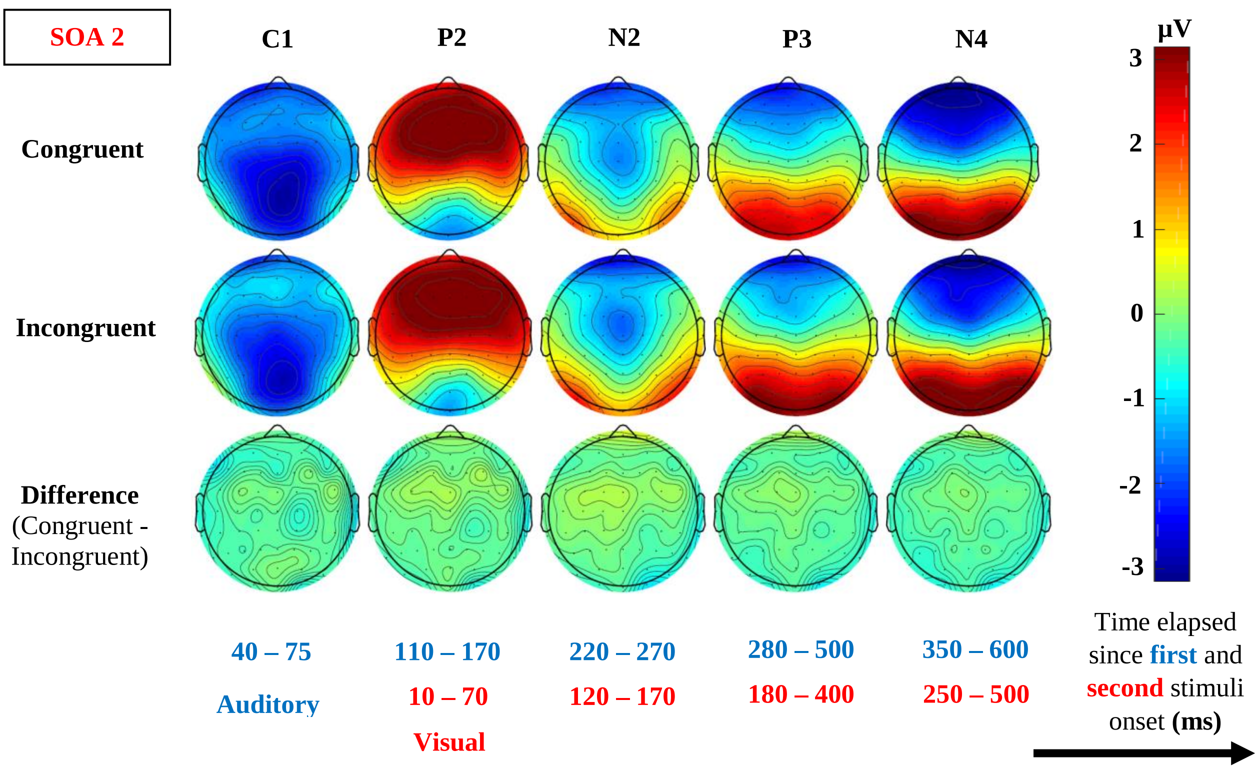

Figure 7b: Experiment 1; Grand average topographic maps for *SOA 2* with congruent, incongruent, and difference (congruent minus incongruent) conditions displayed.

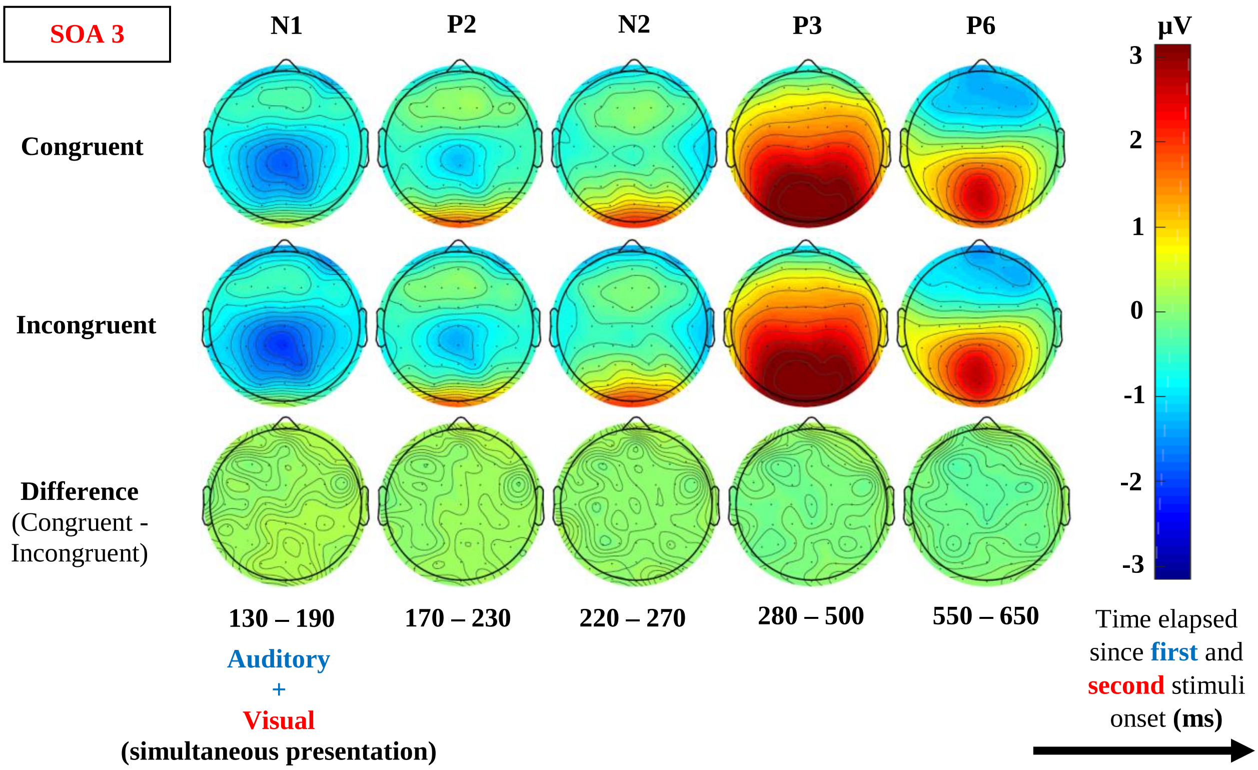

Figure 7c: Experiment 1; grand average topographic maps for *SOA 3* with congruent, incongruent, and difference (congruent minus incongruent) conditions displayed.

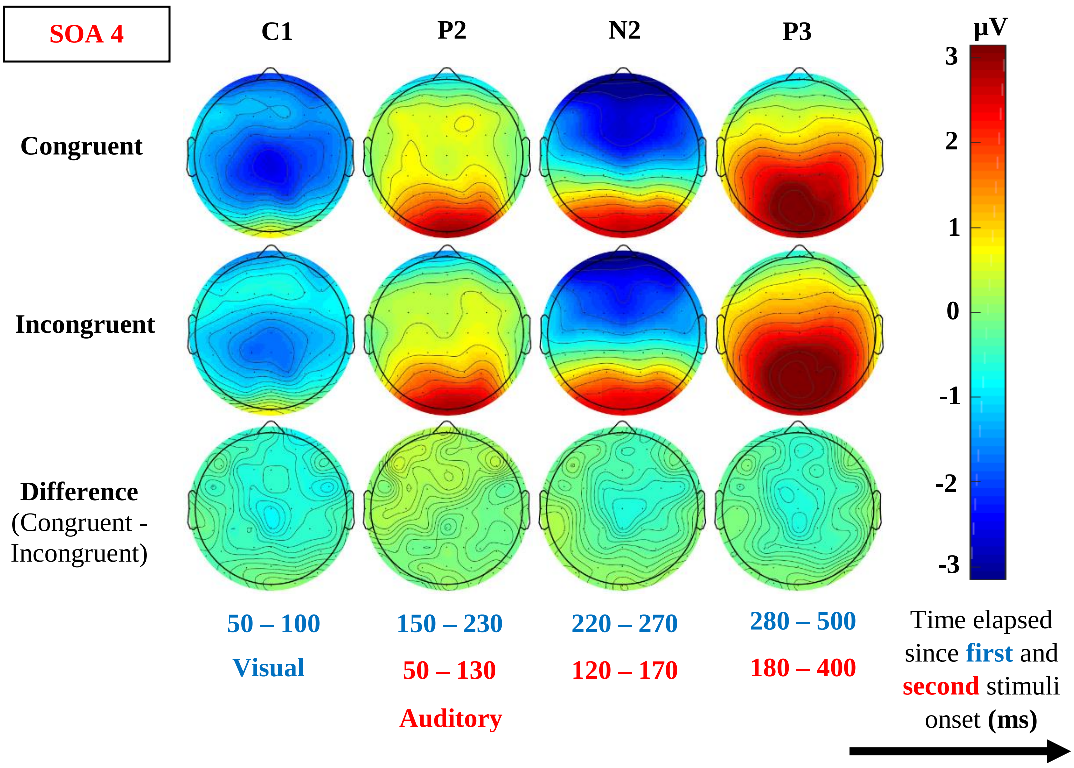

Figure 7d: Experiment 1; grand average topographic maps for *SOA 4* with congruent, incongruent, and difference (congruent minus incongruent) conditions displayed.

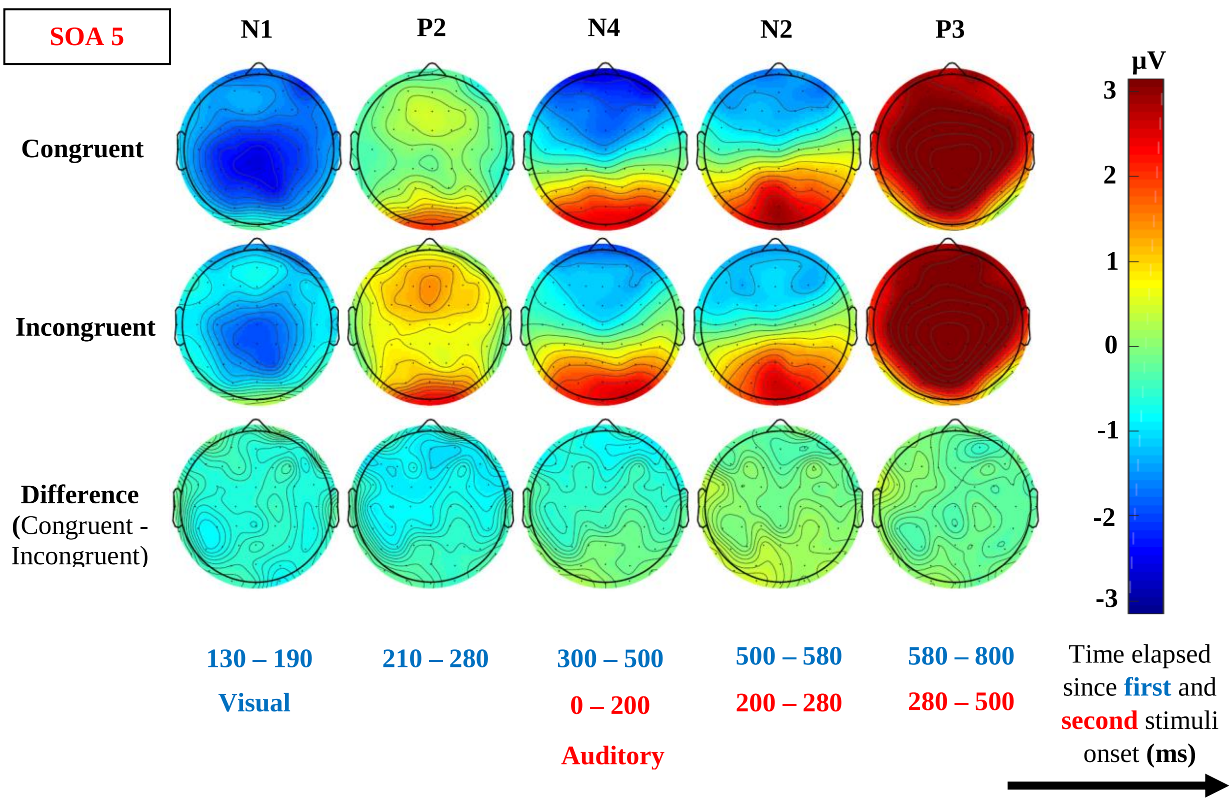

Figure 7e: Experiment 1; grand average topographic maps for *SOA 5* with congruent, incongruent, and difference (congruent minus incongruent) conditions displayed.

Figure 7: Experiment 1; all *five SOAs* displayed as a grand average topographic map of an EEG field in a 2-D circular view (looking down on the transverse plane) using cointerpolation on a fine cartesian grid. Topographic maps are displayed over time regions of interest, corresponding to known visual and auditory ERP components (displayed on top). Because several SOAs consist of an auditory/visual stimulus onset followed by a latency before the second visual/auditory stimulus onset, time regions are labelled underneath for both time elapsed since *first* and *second* stimuli onsets. Corresponding time windows are also overlaid in the ERP responses shown in Figure 5. A colour bar scale is shown on the right-hand side with units of μV.

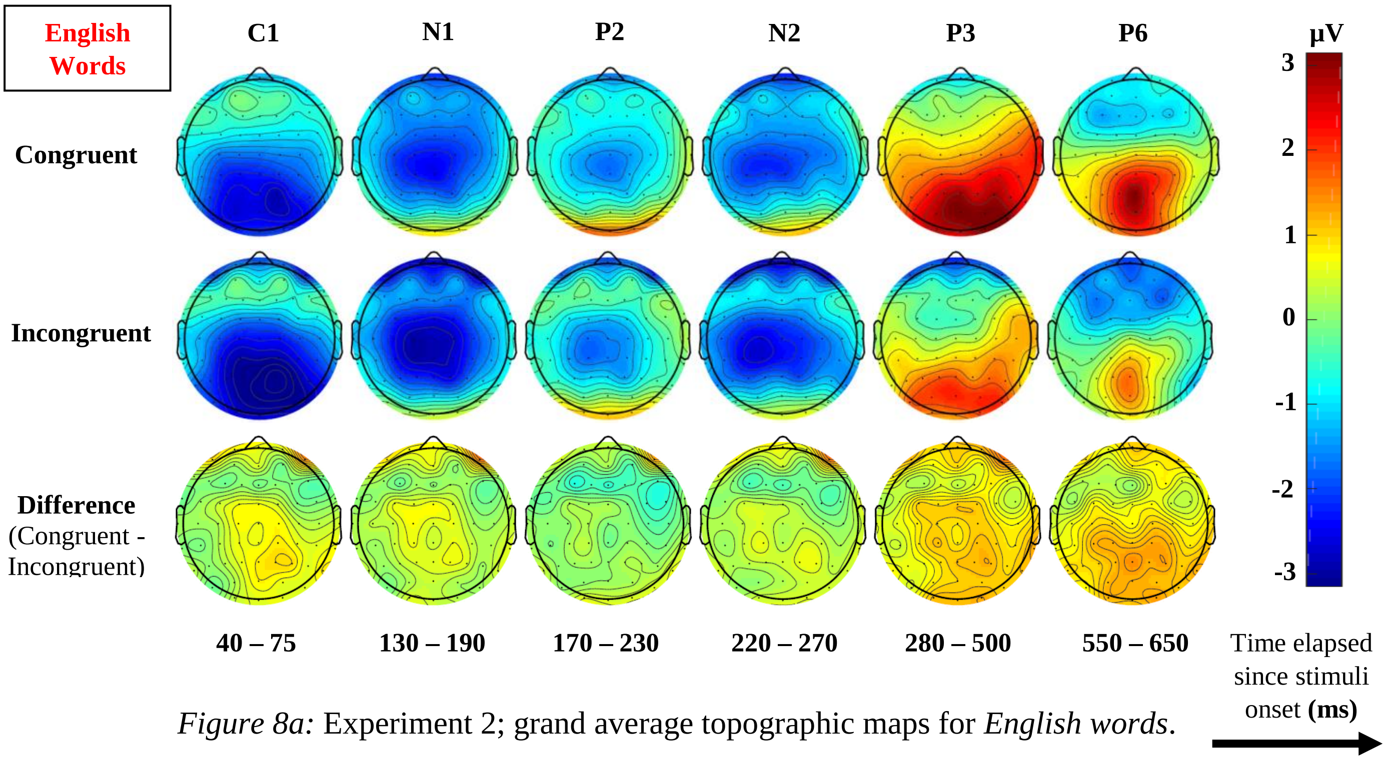

Figure 8a: Experiment 2; grand average topographic maps for *English words*.

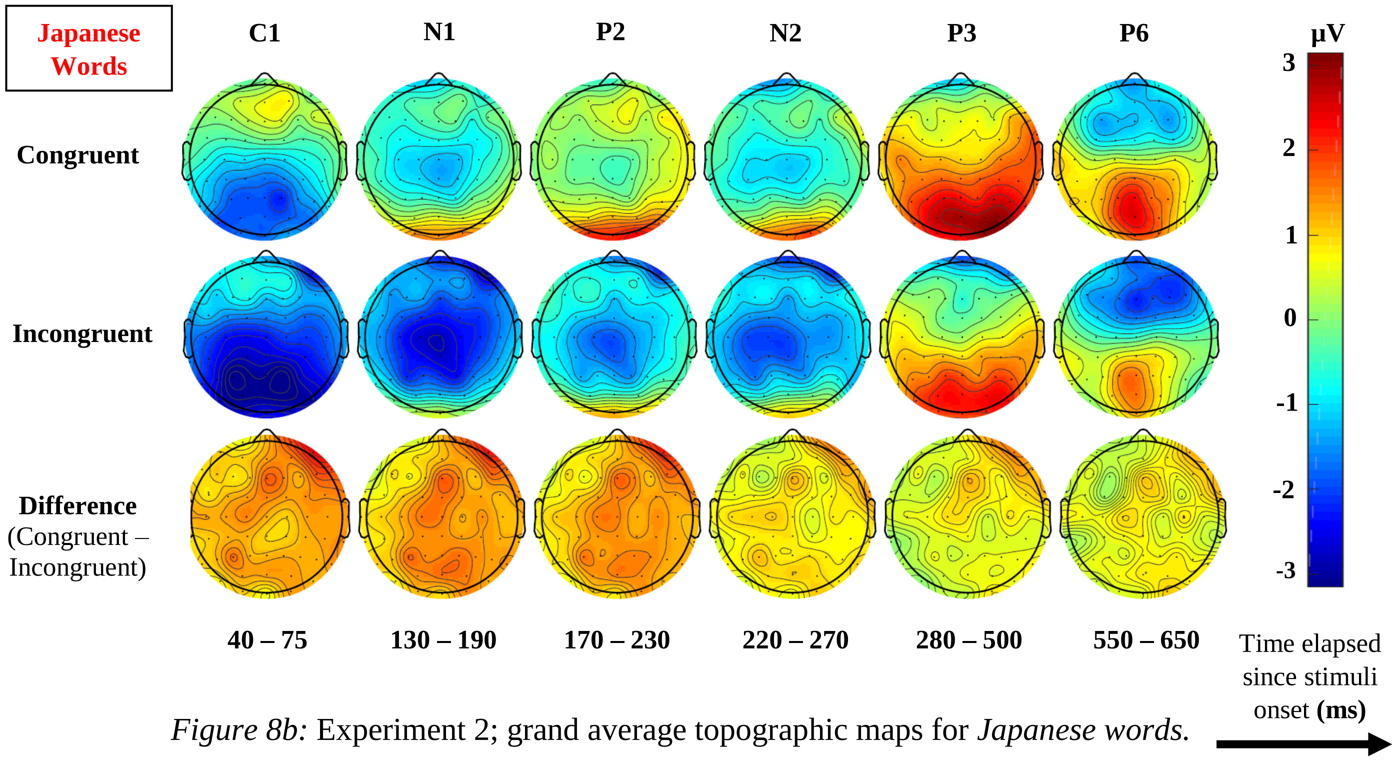

Figure 8b: Experiment 2; grand average topographic maps for *Japanese words.*

Figure 8: Experiment 2; *English* and *Japanese* word conditions displayed as a grand average topographic map of an EEG field in a 2-D circular view (looking down on the transverse plane) using cointerpolation on a fine cartesian grid. Topographic maps are displayed over time regions of interest, corresponding to known visual and auditory ERP components (displayed on top). Corresponding time regions are also overlaid in the ERP responses shown in Figure 6. A colour bar scale is shown on the right-hand side with units of μV.

Discussion

An experiment investigating crossmodal correspondence between auditory and visual modalities was conducted. It consisted of both behavioural and EEG components, and aimed to delve deeper into a reported illusion, whereby auditory stimuli can bias visual motion perception (Maeda et al., 2004), by examining the neural correlates of the effect. It was hypothesised that hearing ascending or descending pitches accompanying an ambiguously moving Gabor patch would influence the participants to perceive motion in a direction congruent to the auditory stimuli. Additionally, it was believed that the semantic content present in the English words “up” and “down”, if presented simultaneously with the same visual stimuli, would bias the individual to perceive congruent visual motion, whereas the Japanese words for “up” and “down” would not. Stimuli were presented over five different SOAs, ranging from -300 to +300 ms, and upon stimuli offset, participants indicated the direction of perceived motion via a 2-AFC task whilst their EEG response was recorded. The resulting behavioural and EEG data for 23 participants were analysed and behaviourally, both hypotheses were confirmed. For the neural data, ERP analysis was performed on 38 ERPs containing seven distinct ERP components (C1, N1, P2, N2, P3, N4, P6), which were plotted in the form of waveform ERP responses and topographic representations. They were statistically analysed using repeated measures ANOVAs to test for any effects of congruency, brain region, and the interaction between them. Six ERP components were found to possess an effect of congruency, five of them belonging to the Japanese word condition and one belonging to the English word condition. Nearly all components displayed an effect of brain region, and one component displayed an interaction effect (P6 from the English Words condition) representing a difference between congruency conditions for frontal and occipital regions.

The results of the behavioural experiment confirm the existence of this type of pitch-induced visual motion perception illusion whilst perhaps hinting at the nature of the underlying crossmodal correspondence. Experiment 1 shows that pitch content alone is enough to slightly bias an individual’s judgement of a moving visual stimuli. It is possible that pitch content of a tone may be a statistically-mediated crossmodal correspondence (Coward & Stevens, 2004; Grassi, 2005), and therefore, we are naturally primed to respond “up” to a visual motion prompt when we are also hear an accompanying pitch glide (Ernst, 2006; Parise & Spence, 2009). This contrasts against Experiment 2, where the salient feature of the auditory stimuli is the semantic information. The results from this experiment clearly show the effect that semantic information can have—an increase in the percentage of congruent decisions of nearly 10%, from 54% in the pitch-based experiment to 64% in the English word-based experiment. The fact that the congruency effect is so much greater for this experiment, perhaps hints that it is mediated by a different factor and that we may be dealing with a semantically mediated crossmodal correspondence (Martino & Marks, 1999). This claim is strengthened by the observation that whilst English stimuli promote a significant bias towards congruent decision making, for Japanese stimuli, participants return congruent decisions at a percentage that is not significantly different from chance. However, it may still possess a statistical component—spectral analysis of the pitch content of the words used in Experiment 2 (see Figure 1) show that when a line of best fit is plotted for the spectrograms of each word, “up” and “down” stimuli also contain average ascending and descending pitches respectively and observed for both languages. Therefore, when the results of Experiments 1 and 2 are considered together, the effect of pitch in Experiment 2 may contribute a small but significant amount to a larger semantic effect.

Moreover, the results from Experiment 1 concerning the SOAs of the stimuli may add to the body of literature investigating the effect that *time* has on crossmodal correspondences. It has been observed that multisensory integration is strongest and most likely to occur for stimuli presented closely in time (Jones & Jarick, 2006; Shore, Barnes, & Spence, 2006; Vroomen & Gelder, 2000). The fact that the zero latency SOA 3 condition was observed to have the greatest congruency effect and that this effect tailed off with increasing stimuli latency gives support to that observation. Furthermore, by considering time in another way, the nature of the crossmodal correspondences could be further explored. For semantic-based auditory stimuli such as the words used in Experiment 2, neural processing and ERP responses happen on a slower timescale than for perceptual responses. Examples of ERPs associated with semantic auditory processing are the N4 and P6 components (Bentin, Kutas, & Hillyard, 1993; Kutas & Federmeier, 2000), which are present especially in frontal areas of the language condition stimuli. These late ERP components may reflect the fact that semantic information is being used at a later decisional stage to influence the participants’ decision making. Therefore, it is possible that the timings of the ERP components for the pitch- and word-based stimuli could be used to argue for either a more decisional or perceptual crossmodal correspondence.

Whilst the present study replicated the main illusion found in the original behavioural experiment (Maeda et al., 2004), the results were not entirely alike for the word conditions. There were similarities in the pitch experiment, discovering that the largest effect of congruency occurred at either zero latency or when the auditory stimuli were presented just after the visual stimuli, which is compatible with the known visual and auditory neural delay (Luce, 1986; Stein & Meredith, 1993). Maeda et al. (2004) found that the Japanese word condition resulted in no significant differences between the English-speaking and Japanese-speaking groups or between the words and chance. For Japanese speakers, the effect was only significant when the words were presented 400 ms after the visual stimuli onset. The present study did replicate the null effect for participants listening to a language they didn't understand, however, a difference in congruency *was* found for the English condition and it was significant at zero latency stimuli presentation. These discrepancies could possibly be caused by the small sample size used in the original study.

In some conditions, the pre-stimulus period was not stable but instead contained a build-up of negative amplitudes, despite randomisation of the first stimuli onset. This may be caused by an expectancy for a certain stimulus to appear (e.g., Britz & Michel, 2010). The sustained build-up of activity is most likely not due to random fluctuations as at very late and very early time epochs, the activity returns to baseline. It is possible that this prediction effect may prime the behavioural response instead of the motion direction decision being influenced by a real perceptual difference. This is unlikely to be the case for trials in Experiment 1, as the difference between congruent and incongruent trials is small, however, for the Japanese word conditions, differences are larger. The difference in pre-stimulus amplitude may account for the near-chance level of decision making present in this condition as what happened after the stimuli onset evidently did not influence the participants' decisions. Larger frontal variations in amplitude can be explained by more frequent muscular activity in the proximity of these electrodes.

The results of the EEG analysis discovered several early ERPs and some late-type ERPs for each condition. However, due to the summation of visual and auditory ERPs, it is difficult to draw concrete conclusions from these analyses. Many of the ERPs identified did not show any effect of congruency which is at odds with the behavioural data, which suggests a significant effect of congruency. Perhaps this is down to signal-to-noise ratio, or because the differences arise from a decisional change which cannot be picked up via ERP analysis. In general, earlier, more perceptual ERPs had higher amplitudes which might reflect the fact that this illusion is more perceptually-driven. Differences between congruent and incongruent trials were in amplitude and not latency, suggesting that the same ERPs are present for both conditions but at varying levels. The first ERP component regularly seen is the negative C1 component, peaking on average between 40 and 75 ms, and strongest for congruent trials in Experiment 1 and incongruent trials in Experiment 2. It is located in occipital and parieto-occipital regions and is associated with automatic visual processing and is not affected by attention (Luck, 2012; Woodman, 2010). It is evoked by a change in luminance, which explains why it is almost always present in our study as a visual stimulus is always presented. The amplitude being highest for incongruent trials in Experiment 2 reflects the fact that incongruent trials have a larger amplitude on average throughout these conditions.

The next ERP component to appear is the N1-P2 complex, which is evoked by any unpredictable auditory (and sometimes visual) stimuli and is involved in perception (Luck, 2012; Näätänen & Picton, 1987). In our data we mainly see the N1 component in frontal areas. The P2 component is present in frontal areas when only a solitary auditory stimulus is presented first (SOAs 1 and 2), and additionally in occipital areas when a solitary visual stimulus is presented first (SOAs 4 and 5) reflecting activity in the visual cortex. The amplitude of the N1-P1 complex is strongly dependent on loudness, arousal, and selective attention. We observe the highest amplitudes for conditions where an auditory stimulus is presented on its own, suggesting that simultaneous presentation and the subsequent multisensory integration may slightly drown out the effect of a single stimulus, and that the visual stimuli may distract attention away from the auditory stimuli. We observe strong amplitudes for congruent Japanese stimuli but not incongruent Japanese stimuli whereas English stimuli show no difference across congruency, compared to incongruent Japanese or congruent English stimuli. If this complex truly is attention mediated, this may suggest that the participant may be more closely attending to the pitch content of the word when no semantic information is available, and the times when they chose a congruent direction are when they have been more closely paying attention to the pitch content of the word. Indeed, it has been previously hypothesised that congruent combinations might yield higher N1/P2 amplitudes (Seo et al., 2010).

Another ERP component consistently appearing is the N2 component, historically associated with mismatch (Sutton, Braren, Zubin, & John, 1965). We find this component located bilaterally in the occipital regions, with no effect of congruency. If the participants had figured out that the visual stimulus was always ambiguous, then this may be an indicator of mismatch between direction. Finally, the P3 component, thought to be evoked in the process of decision making, is found for all trials. However, the only statistically significant effect of congruency for P3 occurs in the English words condition. This suggests a decisional explanation for this type of crossmodal correspondence—semantic priming causes a decision to be made, attributing ambiguous motion to the congruent direction.

There are a few limitations worth mentioning, as well as avenues for further work. Firstly, the fact that Experiment 1 investigated SOA as one of the independent variables brought about a further complication of the summation of auditory and visual ERP components, making it difficult to deconvolve the two. Although SOA 3 and Experiment 2, and the first 300 or 100 ms of the trial remained unaffected by this, future studies should seek to include EEG recording of trials containing only one modality, so that the ERP components from each modality can be more easily attributed. Additionally, reaction time analysis could be beneficial in analysis of the decision stage of the experiments. Speeded 2-AFC tasks could be employed to reduce the amount of decision time available and help to study more perceptual effects by removing any unwanted effect of priming. Future studies could investigate whether words other than those with directional semantic meaning attached (such as “up” and “down”) could elicit the same effect and explore the salience of the pitch content of the word. If other words or sounds can elicit the same effect, would it be due to the spectral features of the stimuli or some other parameter? Furthermore, a pure tone pitch glide constructed using the spectral content of the words “up” and “down” could be used to determine the relative contributions of pitch information and semantics to the observed illusion.

In summary, a behavioural experiment adapted from Maeda et al. (2004) was converted into an EEG experiment in order to better understand the neural substrates of a crossmodal correspondence between pitch and visual motion. Ascending and descending pitches were observed to bias perception of an ambiguous visual stimuli towards the congruent direction 54% of the time, significantly different from the chance level; $t(30) = 4.40$, $p < .001$, with the greatest effect present when the stimuli were presented simultaneously. This effect of congruency increased to 64% when the auditory stimuli were replaced with the spoken words “up” and “down”. The effect disappeared when the words were spoken in Japanese, suggesting the presence of semantically-mediated crossmodal correspondence. ERP results are inconclusive due to the lack of significant effects of congruency, however, the presence of a significant difference in congruency for the P3 component in English word stimuli hints at a decisional process playing a key part in this effect. The relative amplitudes of the N1-P2 component suggests crossmodal stimuli presentation may diminish attention for a single stimulus, possibly affecting the subsequent decision process. Additionally, they may suggest that for word stimuli with no semantic meaning, more attention is paid to the pitch content of the word. These results raise interesting questions about the true nature of this type of crossmodal correspondence and suggest that both a statistically-mediated perceptual effect and a semantically-mediated decisional effect may both play a part in this auditory-visual illusion.

References

Anstis, S., Verstraten, F. A. J., & Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences*, *2*(3), 111–117. https://doi.org/10.1016/S1364-6613(98)01142-5

Belkin, K., Martin, R., Kemp, S. E., & Gilbert, A. N. (1997). Auditory Pitch as a Perceptual Analogue to Odor Quality. *Psychological Science*, *8*(4), 340–342. https://doi.org/10.1111/j.1467-9280.1997.tb00450.x

Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, *30*(2), 161–169. https://doi.org/10.1111/j.1469-8986.1993.tb01729.x

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*(6), 578–584. https://doi.org/10.3758/BF03207374

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663–672. https://doi.org/10.1016/j.neuroimage.2011.06.095

Britz, J., & Michel, C. M. (2010). Errors can be related to pre-stimulus differences in ERP topography and their concomitant sources. *NeuroImage*, *49*(3), 2774–2782. https://doi.org/10.1016/j.neuroimage.2009.10.033

Bronner, K., Frieler, K., Bruhn, H., Hirt, R., & Piper, D. (2008). What is the Sound of Citrus? Research on the Correspondences between the Perception of Sound and Flavour, 7.

Büttner, A. (2017). *Springer Handbook of Odor*. Springer.

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., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649–657. https://doi.org/10.1016/S0960-9822(00)00513-3

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, Validity, and Factor Structure of the Creative Achievement Questionnaire. *Creativity Research Journal*, *17*(1), 37–50. https://doi.org/10.1207/s15326934crj1701_4

Coward, S. W., & Stevens, C. J. (2004). Extracting Meaning from Sound: Nomic Mappings, Everyday Listening, and Perceiving Object Size from Frequency. *The Psychological Record*, *54*(3), 349–364. https://doi.org/10.1007/BF03395478

Crisinel, A.-S., & Spence, C. (2010). A Sweet Sound? Food Names Reveal Implicit Associations between Taste and Pitch. *Perception*, *39*(3), 417–425. https://doi.org/10.1068/p6574

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

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. *Perception*, *131*(Chapter 6). Retrieved from https://pub.uni-bielefeld.de/publication/2355548

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203. https://doi.org/10.3758/BF03193720

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, *67*(2), 274–284. https://doi.org/10.3758/BF03206491

Hornbostel, E. M. (1931). Ueber Geruchshelligkeit. [The distinctness of smell.]. *Pflügers Archiv Für Die Gesamte Physiologie Des Menschen Und Der Tiere*, *228*, 517–537. https://doi.org/10.1007/BF01755351

Jack, C. E., & Thurlow, W. R. (1973). Effects of Degree of Visual Association and Angle of Displacement on the "Ventriloquism" Effect. *Perceptual and Motor Skills*, *37*(3), 967–979. https://doi.org/10.1177/003151257303700360

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.*, *10*, 370–375.

Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time. *Experimental Brain Research*, *174*(3), 588–594. https://doi.org/10.1007/s00221-006-0634-0

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, *416*(6877), 172–174. https://doi.org/10.1038/416172a

Kutas, null, & Federmeier, null. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00213

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press.

Luck, S. J. (2012). *The Oxford handbook of event-related potential components*. New York, NY, US: Oxford University Press.

Ludwig, V. U., & Simner, J. (2013). What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex*, *49*(4), 1089–1099. https://doi.org/10.1016/j.cortex.2012.04.004

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, *14*(23), R990–R991. https://doi.org/10.1016/j.cub.2004.11.018

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, *50*(1), 125–145. https://doi.org/10.1016/S0167-8760(03)00129-6

Marks, L. E., Szczesiul, R., & Ohlott, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 517–534. https://doi.org/10.1037/0096-1523.12.4.517

Martino, G., & Marks, L. E. (1999). Perceptual and Linguistic Interactions in Speeded Classification: Tests of the Semantic Coding Hypothesis. *Perception*, *28*(7), 903–923. https://doi.org/10.1068/p2866

Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: The role of synesthetic correspondence. *Perception*, *29*(6), 745–754. https://doi.org/10.1068/p2984

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and Weak. *Current Directions in Psychological Science*, *10*(2), 61–65. https://doi.org/10.1111/1467-8721.00116

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Parise, C. V., & Spence, C. (2009). ‘When Birds of a Feather Flock Together’: Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes. *PLOS ONE*, *4*(5), e5664. https://doi.org/10.1371/journal.pone.0005664

Sagiv, N., & Ward, J. (2006). Crossmodal interactions: lessons from synesthesia. *Progress in Brain Research*, *155*, 259–271. https://doi.org/10.1016/S0079-6123(06)55015-0

Seo, H. S., Arshamian, A., Schemmer, K., Scheer, I., Sander, T., Ritter, G., & Hummel, T. (2010). Cross-modal integration between odors and abstract symbols. *Neuroscience Letters*, *478*(3), 175–178. https://doi.org/10.1016/j.neulet.2010.05.011

Shore, D. I., Barnes, M. E., & Spence, C. (2006). Temporal aspects of the visuotactile congruency effect. *Neuroscience Letters*, *392*(1–2), 96–100. https://doi.org/10.1016/j.neulet.2005.09.001

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-Visual Crossmodal Integration in Perception of Face Gender. *Current Biology*, *17*(19), 1680–1685. https://doi.org/10.1016/j.cub.2007.08.043

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA, US: The MIT Press.

Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, *18*(1), 70–73. https://doi.org/10.4103/0972-6748.57865

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188.

Thrash, T. M., & Elliot, A. J. (2003). Inspiration as a psychological construct. *Journal of Personality and Social Psychology*, *84*(4), 871–889.

Tootell, R. B., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., … Rosen, B. R. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature*, *375*(6527), 139–141. https://doi.org/10.1038/375139a0

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of Letters and Speech Sounds in the Human Brain. *Neuron*, *43*(2), 271–282. https://doi.org/10.1016/j.neuron.2004.06.025

Vroomen, J., & Gelder, B. de. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1583–1590. https://doi.org/10.1037/0096-1523.26.5.1583

Walker, P., & Smith, S. (1985). Stroop Interference Based on the Multimodal Correlates of Haptic Size and Auditory Pitch. *Perception*, *14*(6), 729–736. https://doi.org/10.1068/p140729

Wallace, M. T., & Stein, B. E. (1996). Sensory organization of the superior colliculus in cat and monkey. *Progress in Brain Research*, *112*, 301–311.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-Colour Synaesthesia: to What Extent Does it Use Cross-Modal Mechanisms Common to us All? *Cortex*, *42*(2), 264–280. https://doi.org/10.1016/S0010-9452(08)70352-6

Woodman, G. F. (2010). A Brief Introduction to the Use of Event-Related Potentials (ERPs) in Studies of Perception and Attention. *Attention, Perception & Psychophysics*, *72*(8). https://doi.org/10.3758/APP.72.8.2031

Yau, J. M., Olenczak, J. B., Dammann, J. F., & Bensmaia, S. J. (2009). Temporal Frequency Channels Are Linked across Audition and Touch. *Current Biology*, *19*(7), 561–566. https://doi.org/10.1016/j.cub.2009.02.013

Appendix

Appendix A1. Written Consent Form Completed by Participants

Consent Form

EEG Study

Participant Name:

Age:

Sex:

Nationality:

Languages Spoken and Proficiency (basic, intermediate, fluent):

Handedness:

Ref. Number:

The participant should complete the whole of this sheet		
	Please tick the appropriate box	
	YES	NO
Are you over 18?		
Do you consent to the use of adhesive stickers?		
Do you consent to the use of conductive gel?		
Do you consent to us recording your EEG?		
Do you consent to us recording medical details provided by you (strictly confidentially)?		
Have you read the Research Participant Information Sheet?		
Do you understand that you will not be referred to by name in any report concerning the study?		
Do you understand that you are free to withdraw from the study:		
• at any time?		
• without having to give a reason for withdrawing?		
Do you agree to take part in this study?		

There are a few medical details which are required prior to participation.

Ticking 'Yes' to these questions doesn't necessarily mean you will not be able to take part.

Please read the following questions and place a tick in the box to indicate your answer. All information given will be treated as confidential.

	Yes	No	Details
Are you currently taking/have you recently taken any prescription or over-the-counter medications? If yes, please give details.			
Have you ever suffered from epilepsy?			
Have you had any surgery in which metal items may have been placed in your head?			
Do you have any history of allergic reactions to skin products, cosmetics or lotions? If yes, please specify.			
Do you have normal or corrected to normal vision?			
Do you have normal hearing?			
Do you have a pacemaker fitted?			
Do you use any other medical electrical device? If yes, please specify.			
Have you been feeling unwell over the last few days? If yes, please give details.			
Do you suffer from any sort of chronic skin condition (dermatitis, eczema, psoriasis etc)? If yes, please specify.			
Have you taken any sort of legal or illegal drug in the past 24 hours? If yes, please give details.			
Have you consumed alcohol in the past 24 hours? If yes, please give details.			
Have you been diagnosed with any kind of psychiatric disorder? If yes, please give details.			
Do you have any family history of psychiatric illness that you know of? If yes, please give details			
Do you have any blood clotting disorder, or are you currently taking any drugs which reduce the effectiveness of blood clotting? If yes, please give details.			

Print Name: **E-Mail:**

Signed: **Date:**

Appendix A2. Debrief Form

Debrief Form

I would like to take this opportunity to say a huge **Thank You** for taking part in my study.

What are we researching?

The completed research will help to gain an understanding of what is happening on a neural level during a crossmodal correspondence between pitch and visual motion.

Information from separate sensory inputs, such as visual, auditory, and tactile stimuli, were once thought to be processed independently in the brain. However, it has since been shown that sensory information from different modalities can interact on a neural level, leading to changes in perception and cognition; this is known as crossmodal correspondence.

This study seeks to investigate the correspondences between auditory-visual stimuli, specifically, auditory sounds moving upwards/downwards in pitch and a simultaneously presented moving visual stimuli. In 2004, a paper by Maeda et al. investigated how perception could be affected by a crossmodal correspondence between pitch and visual motion—showing that participants simultaneously presented with an ascending/descending sound and ambiguous visual motion are significantly more likely to perceive the motion as travelling in the same direction as the sound. Our research is attempting to explore what is happening on a neural level during this phenomenon.

Please be assured, all data collected will be treated in the strictest confidence. You are free to withdraw your data from the research at any time by contacting the primary researcher (alasc001@gold.ac.uk) or MSc project supervisor (j.bhattacharya@gold.ac.uk).

If you were unduly or unexpectedly affected by taking part in the study, please don't hesitate to provide the researcher with feedback. If you feel unable to talk with the researcher, then please either contact the supervisor or the Head of Psychology.

If you have any friends who might be interested in taking part and earning some money, please ask them to contact me at alasc001@gold.ac.uk or 07960704244. Thank you!

Appendix A3. Information Sheet

Participant Information Sheet

MSc EEG Study

Welcome!

You are being invited to take part in a student research project. It is important that you familiarise yourself with what the experiment involves before you decide to participate. Please read the following information carefully and feel free to ask any questions if anything is not clear.

Purpose of this study

The purpose of this project is to further our understanding of how the brain functions when different stimuli are presented to you. More information can be given in the debrief following the experiment. Activity will be measured using electroencephalography (EEG) which involves placing small electrodes on the scalp to passively record electrical activity from the brain.

Do I have to take part?

Taking part in this study is completely voluntary. Should you decide to participate, you will be given this information sheet to keep and be asked to sign a consent form. You will be free to withdraw from this study at any time and will not be required to give a reason. There are no repercussions for withdrawing from this study.

What will happen to me if I take part?

You will attend one 3-hour EEG session after which you will receive a small monetary compensation for your time. The session will take place on the Goldsmiths College campus on the 2nd floor of the Ben Pimlott Building.

You will be required to wear an EEG cap, allowing the placement of electrodes onto your scalp. These electrodes will require the application of conductive gel to the scalp to ensure a strong signal is obtained. The gel will be applied to the surface of your scalp using a syringe. Facilities to wash your hair after the EEG session will be provided.

Once the EEG has been set up, you will be asked to take a seat in the testing room, in front of a computer screen. Images will be presented on the screen along with sounds played through a speaker. You will be expected to respond to the stimuli as accurately and quickly as possible.

What do I have to do?

As we are measuring brain activity, there are a few important lifestyle restrictions that will need to be adhered to before attending.

1) Do not take any depressants/sedatives/relaxers (e.g., alcohol, marijuana or other recreational mind-altering drugs) within 24 hours of your EEG.

2) Do not take any stimulants (e.g., caffeine, cigarettes, tea, coffee, chocolate, energy drinks) within 8 hours of your EEG session.

3) Be sure to drink plenty of water in the 24 hours prior to your EEG session.

4) Please aim to get a good night's sleep (8 hours) before the day of your EEG session. However, do not take any prescription or over the counter sleeping aids. If you have any problem with sleep or are exceptionally tired on the day of your EEG, please tell the researcher on the day.

5) Please ensure you eat a good meal beforehand so that you will not be hungry during the EEG session. However, try to avoid foods/drinks that are high in sugar or fat.

6) On the day of your EEG, please wash your hair thoroughly with shampoo, focussing on your scalp. Rinse your hair thoroughly to remove any shampoo from the scalp. Do not use conditioners, gels, hairsprays or other hair products after cleaning your hair as this may make it difficult to establish a strong signal for the EEG.

What are the possible disadvantages and risks of taking part?

EEG is not associated with any physiological or psychological risks. Whilst currently no risk is associated with EEG, if you feel that you're unwell or unfit to participate at any point then please let the researcher know so that proper health and safety guidelines may be followed and the appropriate authorities may be informed if needed.

As the experiment will take place around electrical equipment within a lab environment; in the unlikely event of a fire, participants and colleagues will be evacuated to the nearest available exit and escorted to an assembly point. Proper authorities will then be notified.

What are the possible benefits of taking part?

There are no direct benefits of taking part in this study besides monetary reimbursement and the potential novel experience of EEG brain recording. However, through participation you will be aiding in scientific research—helping to unravel the mysteries surrounding the human brain.

What if something goes wrong?

As previously stated, there is no direct risk in participating in this study. However, in the very unlikely event that you are harmed during participation, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for legal action but may have to fund it yourself.

Will my taking part in this study be kept confidential?

All information which is collected about you during this study will be kept strictly confidential. Any information about you which leaves the University will not contain your name and address so that you cannot be identified.

What will happen to the results of the research study?

The results of the research will be presented to the members of the psychology department at Goldsmiths and may be published in a research journal. A copy of the results may be obtained from the researcher, Alex Lascelles. Any personal information will be omitted from the write-up to ensure confidentiality.

Who has reviewed the study?

This study has been reviewed and approved by the Goldsmiths Research Ethics Committee.

Research Integrity

Goldsmiths, University of London, is committed to compliance with the Universities UK Research Integrity Concordat. You are entitled to expect the highest level of integrity from our researchers during the course of their research.

Contacts for Further Information

If you decide to participate you will be given a copy of this information sheet and a debrief form on completion of the experiment. However, should you require further information please feel free to contact us:

Researcher: Mr Alex Lascelles
E-mail: alasc001@gold.ac.uk
Phone: 0796 0704 244

Supervisor: Prof Joydeep Bhattacharya
Email: j.bhattacharya@gold.ac.uk

Department of Psychology

Goldsmiths, University of London
Whitehead Building

New Cross, SE14 6NW

Phone: +44 20 7919 7870 / 7871

Thank you very much for taking part in this study!

Appendix B1. Inspiration Questionnaire

Below are four statements, each followed by two questions. The questions concern how often and how deeply/strongly you experience what is described in the statement. Please answer both questions after each statement by circling numbers from 1 to 7.

1. I experience inspiration.

1a. How often does this happen?	1 never	2	3	4	5	6	7 very often
1b. How deeply or strongly (in general)?	1 not at all	2	3	4	5	6	7 very

2. Something I encounter, or experience inspires me.

2a. How often does this happen?	1 never	2	3	4	5	6	7 very often
2b. How deeply or strongly (in general)?	1 not at all	2	3	4	5	6	7 very

3. I am inspired to do something.

3a. How often does this happen?	1 never	2	3	4	5	6	7 very often
3b. How deeply or strongly (in general)?	1 not at all	2	3	4	5	6	7 very

4. I feel inspired.

4a. How often does this happen?	1 never	2	3	4	5	6	7 very often
4b. How deeply or strongly (in general)?	1 not at all	2	3	4	5	6	7 very

Key for the Inspiration Scale:
Inspiration frequency subscale: sum of items 1a, 2a, 3a, 4a
Inspiration intensity subscale: sum of items 1b, 2b, 3b, 4b
Overall scale: sum of items 1a, 1b, 2a, 2b, 3a, 3b, 4a, 4b

Reference:

Thrash & Elliot, 2003. Inspiration as a psychological construct. Journal of Personality and Social Psychology.

Appendix B2. Ten-Item Personality Inventory

Ten-Item Personality Inventory-(TIPI)

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree strongly (1) Disagree moderately (2) Disagree a little (3) Neither agree nor disagree (4) Agree a little (5) Agree moderately (6) Agree strongly (7)

I see myself as:
1. ____ Extraverted, enthusiastic.
2. ____ Critical, quarrelsome.
3. ____ Dependable, self-disciplined.
4. ____ Anxious, easily upset.
5. ____ Open to new experiences, complex.
6. ____ Reserved, quiet.
7. ____ Sympathetic, warm.
8. ____ Disorganized, careless.
9. ____ Calm, emotionally stable.
10. ____ Conventional, uncreative.

Reference:
Gosling, Rentfrow, & Swann, 2003. Journal of Research in Personality 37 (2003) 504–528

Appendix B3. Creative Achievement Questionnaire

Creative Achievement Questionnaire
Shelley Carson
Harvard University

I. *Place a check mark beside the areas in which you feel you have more talent, ability, or training than the average person.*

__ visual arts (painting, sculpture)
__ music
__ dance
__ individual sports (tennis, golf)
__ team sports
__ architectural design
__ entrepreneurial ventures
__ creative writing
__ humor
__ inventions
__ scientific inquiry
__ theater and film
__ culinary arts

II. *Place a check mark beside sentences that apply to you. Next to sentences with an asterisk (*), write the number of times this sentence applies to you.*

A. Visual Arts (painting, sculpture)
__ 0. I have no training or recognized talent in this area. (Skip to Music).
__1. I have taken lessons in this area.
__2. People have commented on my talent in this area.
__3. I have won a prize or prizes at a juried art show.
__4. I have had a showing of my work in a gallery.
__5. I have sold a piece of my work.
__6. My work has been critiqued in local publications.
*__7. My work has been critiqued in national publications.

B. Music
__0. I have no training or recognized talent in this area (Skip to Dance).
__1. I play one or more musical instruments proficiently.
__2. I have played with a recognized orchestra or band.
__3. I have composed an original piece of music.
__4. My musical talent has been critiqued in a local publication.
__5. My composition has been recorded.
__6. Recordings of my composition have been sold publicly.
*__7. My compositions have been critiqued in a national publication.

C. Dance
__0. I have no training or recognized talent in this area (Skip to Architecture)
__1. I have danced with a recognized dance company.

__2. I have choreographed an original dance number.
__3. My choreography has been performed publicly.
__4. My dance abilities have been critiqued in a local publication.
__5. I have choreographed dance professionally.
__6. My choreography has been recognized by a local publication.
__7. My choreography has been recognized by a national publication.

D. Architectural Design
__0. I do not have training or recognized talent in this area (Skip to Writing).
__1. I have designed an original structure.
__2. A structure designed by me has been constructed.
__3. I have sold an original architectural design.
__4. A structure that I have designed and sold has been built professionally.
__5. My architectural design has won an award or awards.
__ 6. My architectural design has been recognized in a local publication.
*__7. My architectural design has been recognized in a national publication.

E. Creative Writing
__0. I do not have training or recognized talent in this area (Skip to Humor).
__1. I have written an original short work (poem or short story).
__2. My work has won an award or prize.
__3. I have written an original long work (epic, novel, or play).
__4. I have sold my work to a publisher.
__5. My work has been printed and sold publicly.
__6. My work has been reviewed in local publications.
*__7. My work has been reviewed in national publications.

F. Humor
__0. I do not have recognized talent in this area (Skip to Inventions).
__1. People have often commented on my original sense of humor.
__2. I have created jokes that are now regularly repeated by others.
__3. I have written jokes for other people.
__ 4. I have written a joke or cartoon that has been published.
__5. I have worked as a professional comedian.
__6. I have worked as a professional comedy writer.
__7. My humor has been recognized in a national publication.

G. Inventions
__0. I do not have recognized talent in this area.
__1. I regularly find novel uses for household objects.
__2. I have sketched out an invention and worked on its design flaws.
__3. I have created original software for a computer.
__4. I have built a prototype of one of my designed inventions.
__5. I have sold one of my inventions to people I know.
*__6. I have received a patent for one of my inventions.
*__7. I have sold one of my inventions to a manufacturing firm.

H. Scientific Discovery
__0. I do not have training or recognized ability in this field (Skip to Theater)
__1. I often think about ways that scientific problems could be solved.
__2. I have won a prize at a science fair or other local competition.
__3. I have received a scholarship based on my work in science or medicine.
__4. I have been author or co-author of a study published in a scientific journal.
*__5. I have won a national prize in the field of science or medicine.
*__6. I have received a grant to pursue my work in science or medicine.
__7. My work has been cited by other scientists in national publications.

I. Theater and Film
__0. I do not have training or recognized ability in this field.
__1. I have performed in theater or film.
__2. My acting abilities have been recognized in a local publication.
__3. I have directed or produced a theater or film production.
__4. I have won an award or prize for acting in theater or film.
__5. I have been paid to act in theater or film.
__6. I have been paid to direct a theater or film production.
*__7. My theatrical work has been recognized in a national publication.

J. Culinary Arts
__0. I do not have training or experience in this field.
__1. I often experiment with recipes.
__2. My recipes have been published in a local cookook.
__3. My recipes have been used in restaurants or other public venues.
__4. I have been asked to prepare food for celebrities or dignitaries.
__5. My recipes have won a prize or award.
__6. I have received a degree in culinary arts.
*__7. My recipes have been published nationally.

K. Please list other creative achievements not mentioned above.

Scoring of the Creative Achievement Questionnaire

1. Each check marked item receives the number of points represented by the question number adjacent to the checkmark.
2. If an item is marked by an asterisk, multiply the number of times the item has been achieved by the number of the question to determine points for that item.
3. Sum the total number of points within each do- main to determine the domain score.
4. Sum all ten domain scores to determine the total CAQ score.

Reference:
Carson, Peterson, & Higgins, 2005. Reliability, Validity, and Factor Structure of the Creative Achievement Questionnaire

Appendix C. ERP Statistics

Table 4

Repeated Measures ANOVA Results for ERP Analysis of Experiments 1 & 2

Condition	ERP Component	Comparison	Time-Window (ms)	*F*	*df*	df_{err}	*p*	η^2
SOA 1	C1	Congruency	-260 to -225	1.30	1	22	.266	.056
		Region		11.7	2	44	***	.347
		Congruency*Region		.165	2	44	.849	.007
	N1	Congruency	-170 to -110	1.29	1	22	.268	.055
		Region		17.5	2	44	***	.444
		Congruency*Region		1.38	2	44	.262	.059
	P2	Congruency	-130 to -70	.348	1	22	.561	.016
		Region		18.8	2	44	***	.460
		Congruency*Region		1.08	2	44	.348	.047
	N2	Congruency	-80 to -30	.015	1	22	.902	.001
		Region		22.2	2	44	***	.503
		Congruency*Region		2.49	2	44	.094	.102
	N1	Congruency	130 to 190	2.16	1	22	.156	.089
		Region		32.0	2	44	***	.592
		Congruency*Region		1.26	2	44	.294	.054
	P2	Congruency	170 to 230	2.02	1	22	.169	.084
		Region		34.9	2	44	***	.614
		Congruency*Region		.316	2	44	.731	.014
	P3	Congruency	280 to 500	1.53	1	22	.229	.065
		Region		23.6	2	44	***	.518
		Congruency*Region		.875	2	44	.424	.038
SOA 2	C1	Congruency	-60 to -25	1.06	1	22	.315	.046
		Region		21.9	2	44	***	.499
		Congruency*Region		1.01	2	44	.374	.044
	P2	Congruency	10 to 70	.289	1	22	.596	.013
		Region		24.6	2	44	***	.528
		Congruency*Region		3.02	2	44	.059	.121

	N2	Congruency	120 to 170	.195	1	22	.663	.009
		Region		23.2	2	44	***	.513
		Congruency*Region		.772	2	44	.468	.034
	P3	Congruency	180 to 400	.778	1	22	.387	.034
		Region		21.7	2	44	***	.496
		Congruency*Region		.471	2	44	.628	.021
	N4	Congruency	250 to 500	1.14	1	22	.297	.049
		Region		5.67	2	44	**	.205
		Congruency*Region		.323	2	44	.726	.014
SOA 3	N1	Congruency	130 to 190	1.73	1	22	.202	.073
		Region		.194	2	44	.824	.009
		Congruency*Region		.336	2	44	.716	.015
	P2	Congruency	170 to 230	1.62	1	22	.216	.069
		Region		5.63	2	44	**	.204
		Congruency*Region		.040	2	44	.961	.002
	N2	Congruency	220 to 270	1.06	1	22	.314	.046
		Region		4.77	2	44	.013	.178
		Congruency*Region		.031	2	44	.969	.001
	P3	Congruency	280 to 500	.098	1	22	.757	.004
		Region		12.3	2	44	***	.358
		Congruency*Region		.264	2	44	.769	.012
	P6	Congruency	550 to 650	.013	1	22	.911	.001
		Region		11.8	2	44	***	.349
		Congruency*Region		1.36	2	44	.258	.058
SOA 4	C1	Congruency	150 to 200	.680	1	22	.416	.030
		Region		5.57	2	44	.007	.202
		Congruency*Region		2.21	2	44	.122	.091
	P2	Congruency	250 to 330	.033	1	22	.857	.002
		Region		8.47	2	44	***	.278
		Congruency*Region		.680	2	44	.512	.030
	N2	Congruency	320 to 370	.111	1	22	.742	.005
		Region		31.7	2	44	***	.591
		Congruency*Region		2.87	2	44	.068	.115
	P3	Congruency	380 to 600	.375	1	22	.547	.017

		Region		8.95	2	44	***	.289
		Congruency*Region		2.14	2	44	.130	.089
SOA 5	N1	Congruency	130 to 190	1.51	1	22	.231	.064
		Region		2.56	2	44	.089	.104
		Congruency*Region		.431	2	44	.652	.019
	P2	Congruency	210 to 280	2.62	1	22	.147	.093
		Region		3.75	2	44	*	.146
		Congruency*Region		1.46	2	44	.244	.062
	N4	Congruency	300 to 500	.563	1	22	.461	.025
		Region		20.8	2	44	***	.486
		Congruency*Region		2.91	2	44	.065	.117
	N2	Congruency	500 to 580	.069	1	22	.795	.003
		Region		17.6	2	44	***	.444
		Congruency*Region		1.40	2	44	.258	.060
	P3	Congruency	550 to 800	.068	1	22	.796	.003
		Region		4.43	2	44	*	.168
		Congruency*Region		.503	2	44	.608	.022
English Words	C1	Congruency	40 to 75	.513	1	22	.481	.023
		Region		20.5	2	44	***	.483
		Congruency*Region		.325	2	44	.694	.015
	N1	Congruency	130 to 190	.563	1	22	.461	.025
		Region		12.3	2	44	***	.358
		Congruency*Region		.015	2	44	.985	.001
	P2	Congruency	170 to 230	.033	1	22	.858	.001
		Region		3.89	2	44	.028	.150
		Congruency*Region		.707	2	44	.499	.031
	N2	Congruency	220 to 270	.418	1	22	.525	.019
		Region		2.71	2	44	.078	.110
		Congruency*Region		.396	2	44	.675	.018
	P3	Congruency	280 to 500	5.88	1	22	.024	.211
		Region		9.00	2	44	***	.290
		Congruency*Region		.492	2	44	.615	.022
	P6	Congruency	550 to 650	3.20	1	22	.087	.127
		Region		12.7	2	44	***	.366

		Congruency*Region		2.71	2	44	***	.110
Japanese Words	C1	Congruency	40 to 75	11.3	1	22	**	.338
		Region		11.1	2	44	***	.336
		Congruency*Region		.132	2	44	.877	.006
	N1	Congruency	130 to 190	14.8	1	22	***	.402
		Region		8.28	2	44	***	.273
		Congruency*Region		.581	2	44	.563	.026
	P2	Congruency	170 to 230	11.8	1	22	**	.349
		Region		4.62	2	44	*	.174
		Congruency*Region		.530	2	44	.592	.024
	N2	Congruency	220 to 270	6.80	1	22	**	.236
		Region		2.82	2	44	.070	.114
		Congruency*Region		.010	2	44	.990	.000
	P3	Congruency	280 to 500	4.43	1	22	*	.167
		Region		7.26	2	44	**	.248
		Congruency*Region		.667	2	44	.518	.029
	P6	Congruency	550 to 650	3.05	1	22	.095	.122
		Region		14.4	2	44	**	.395
		Congruency*Region		.266	2	44	.768	.012

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$.

Appendix D1. Participant Information and Response Table (See attached data submission)

Appendix D2. Stimuli (See attached data submission)

Appendix D3. EEG MATLAB Scripts (See attached data submission)

Appendix D4. Raw EEG data (See attached data submission)