Appendix A

Adult EEG Study: Spatially Congruent and Incongruent Audiovisual Interactions

A.1 Overview

Most detailed studies on multisensory integration at the neuronal level have focused on multisensory neurons in the mammalian superior colliculus. These studies have revealed limits on the spatial and temporal disparities between the signal sources of multiple modalities that can be tolerated while still producing a supra-additive enhanced firing rate, as well as larger disparities within which a depressed firing rate is produced (Stein & Meredith, 1993). However, these principles of spatiotemporal integration also hold for multisensory neurons in other areas of the brain, including areas that are mediated by the superior colliculus (Stein, Hunneycutt, & Meredith, 1988) and more globally in the neocortex (Wallace & Stein, 1996). So, while behavioral data can reveal many interesting task and stimulus dependent effects of spatial and temporal congruency—including improved response latencies to spatially congruent targets (Hughes, Nelson, & Aronchick, 1998; Hughes et al., 1994) and slowed response latencies to spatially incongruent targets (Hairston et al., 2003; Harrington & Peck, 1998)—the critical question is what areas of the brain are responsible for these behavioral outcomes, in what order are they recruited after the signals arrive at the sensory receptors, and how does either the areas involved or the timing differ in the developing brain? The high-temporal resolution of electroencephalogram (EEG) recordings, as well as their non-invasive nature, make it suitable for investigating the time course and brain areas involved in multisensory processing in the human brain, and ideal for examining the infant brain. Given that even among adult multisensory studies, minor variations in stimulus type and intensity, task, or procedure used can dramatically influence the results, before attempting to record from a developing brain it is necessary to have an adult baseline to provide the proper perspective.

Multi-channel event-related brain potentials (ERPs) and response latencies were recorded from adult human subjects while performing saccades to unimodal auditory (A) and visual (V), as well as spatially congruent (AV_C) or spatially incongruent (AV_I) audiovisual stimuli at either $\pm 25°$ from midline. Subjects were instructed to perform a saccade as quickly as possible to the target location (the visual component in the spatially incongruent condition). The audiovisual interaction was assessed by comparing the ERPs of the two audiovisual stimuli to the sum of the unimodal visual and auditory ERPs (AV_C versus $A\&V$, AV_I versus $A\&V$); the effect of congruency was assessed by a comparison between the summed ERPs of the spatially congruent and the spatially incongruent trials (AV_C versus AV_I). For AV_C, the earliest audiovisual interaction was found in the right centro-parietal regions at 70 ms (post-stimulus onset), with a more robust, stronger and persistent interaction—beginning at 146 ms—found consistently over many brain regions. However, only a very weak audiovisual interaction in the right frontal region was found for AV_I. Comparisons between the event-related potentials for spatially congruent and incongruent audiovisual stimuli showed robust differences over parietal and frontal cortical regions as early as 150 ms after stimulus onset. Consistent with previous behavioral studies (Study 1 and Study 2), analysis of subjects overt response showed a significantly quicker latency toward AV_C compared to all other stimulus conditions.

A.2 Experimental Design and Methods

A.2.1 Participants

Eleven healthy adults (eight naïve) participated in this study after giving written informed consent. All were free from neurological disease, had normal hearing ability, and normal or corrected-to-normal vision.

A.2.2 Apparatus and Stimuli

The experiment apparatus and the stimuli were identical to those used in the three preceding infant studies. Four different target modalities were presented, at either $\pm 25°$: auditory-only (A), visual-only (V), spatially congruent audiovisual (AV_C), and spatially incongruent audiovisual ($AV_I, 50°$ disparity) stimuli, producing eight target types (modality x position).

A.2.3 Procedure

Procedures for trial presentation were the same as the preceding studies, though of much greater duration given the longer attention span of adult subjects, and with all modality conditions (unimodal, spatially congruent, and spatially incongruent) incorporated in to one experiment. One hundred blocks were presented for each of the eight target types (A, V, AV_C, and AV_I at $\pm 25°$), for a total of 800 trials. Brief breaks ($\sim$ 5 minutes) were permitted every 200 trials, as needed. Subjects had their heads stabilized with a chin-rest and were instructed to perform a saccade in the direction of the target as quickly as possible. In addition to the use of digital video data to capture and calculate response latency onsets (same as before), continuous EEG signals were acquired from BioSemi ActiveTwo electrodes with built-in DC coupled amplifiers (BioSemi, Netherlands) at 28 locations spaced over the scalp according to the standard 10–20 locations. Eye movements and eye blinks were measured separately by two additional electrodes placed at the left and right external canthi. Signals were band-pass filtered (0.1–100 Hz) and digitized at a rate of 512 Hz. All scalp electrodes were algebraically re-referenced with respect to the average signals from both ear lobes. The criteria used to discard trials was based upon the behavioral data (response latencies < 100 ms, blinks during target onset, or a response in the wrong direction), as well as trials where the scalp potential exceeded 70 V. Artifact-free EEG epochs (100 ms before stimulus onset, and 500 ms post stimulus onset) were averaged off-line separately for each of the target types, to produce the associated ERP response, and digitally band-pass (0.1–30 Hz) filtered. All good trials were pooled across subjects to calculate the grand mean.

A.3 Results

A.3.1 Audiovisual Interaction: Behavioral Data

A 1-way univariate ANOVA (4) was performed with stimulus modality (A, V, AV_C, and AV_I) as the within-subject factor, on the response latency data, and a main effect was found ($p < 0.001$). Consistent with results from Study 1 and Study 2, this main effect was exclusively due to the much faster response latency toward spatially congruent targets (AV_C) than either unimodal or the spatially incongruent (AV_I) targets, which were not significantly different from each other (Figure A.1). A test of the Race Model Inequality (not shown) was performed for both AV_C and AV_I and, as expected, we only found a violation—indicating non-linear neural summation—for the spatially congruent (AV_C) condition, but not the spatially incongruent (AV_I) condition.

A.3.2 Audiovisual Interaction: Event-Related Potentials

A.3.2.1 Bimodal versus Unimodal ERPs

In order to begin our analysis of the time course and the brain areas important in audiovisual integration, we calculated the differential event-related potential (ERP) response for both bimodal conditions, AV_C and AV_I, against the sum of the unimodal ERPs:

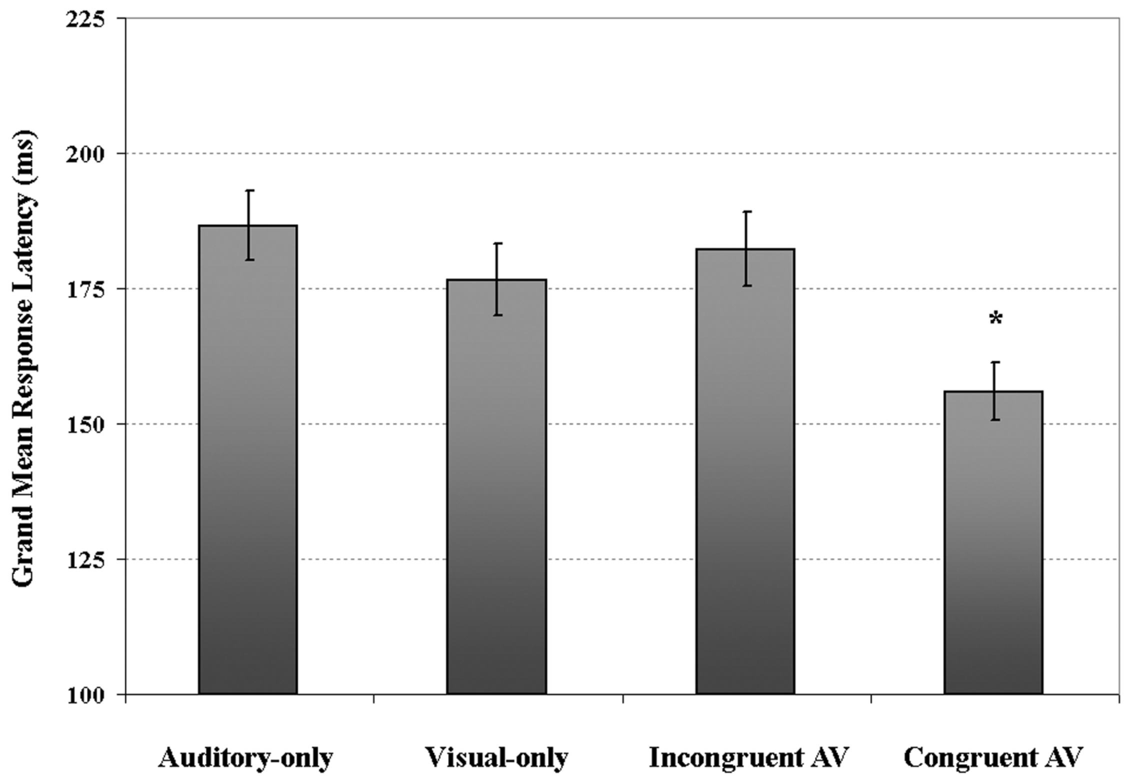

Figure A.1: Main effect of modality on response latency in normal adults. Error bars represent the standard error of the mean.

$$ERP_t(A_i \otimes V_j) = ERP_t(A_iV_j) - [ERP_t(A_i) + ERP_t(V_j)]$$

where $(A \otimes V)$ is a measure of the interaction effect between synchronously presented inputs and i and j represent the absolute position ($\pm 25°$; $i = j$ for AV_C and $i \neq j$ for AV_I) of the auditory and visual components at each time point t since target onset. These differential ERPs were calculated at each scalp electrode and statistically compared against zero voltage (i.e. no interaction) using a point-wise running t-test. A significant audiovisual interaction was defined as there being at least 10 successive data points—spanning approximately 20 ms—meeting the statistically significant criteria ($p < 0.01$). This criterion was found (Guthrie & Buchwald, 1991) to be more stringent than conventional Bonferroni correction for multiple comparison, thus making it appropriate when a large number of t-tests are calculated across many electrodes and epochs (Molholm et al., 2002). Depending on the sign (either positive or negative) of the differential response, two types of interaction—supra- or sub-additive—can be distinguished.

Plotting all significant spatially congruent audiovisual interactions (difference between AV_C and sum of unimodal ERPs) with a minimum duration of 20 ms for all channels (Figure A.2a), we found the earliest audiovisual interaction 70–104 ms after target onset, at the three electrodes in the right centro-parietal regions (C4, CP2, CP6) and one electrode in the frontal region (Fz) of the scalp. A more robust audiovisual interaction began 146 ms post-onset over a broad range of electrodes in the right hemisphere, with a secondary interaction onset in the left fronto-central region at 177 ms post-onset. These latter interactions were found to be robust over a long duration of time (~ 200 ms); the most consistent and long-lasting interactions in the right fronto-central regions.

In contrast, a similar plot of significant spatially incongruent audiovisual interactions (difference between AV_I and sum of unimodal) was also made, showing a much different picture (Figure A.2b). Unlike the spatially congruent condition, there was a much weaker and scattered profile of audiovisual interaction. The earliest interaction was again found 70–112 ms after target onset, but was an isolated activation in the mid-central (Cz) electrode, followed shortly thereafter in the mid-frontal (Fz) and left temporal (T7) electrodes, and then the right frontal (F8) and central (C4) electrodes. However, no conspicuous and consistent audiovisual interaction was found in the later period ($>$ 140 ms) for the majority of the electrode regions. Only three electrodes showed any effect at all: left frontal (F3) at 176 ms, and mid-central and mid-frontal (Cz, Fz) returning at $>$ 228 ms—the latter two electrode onsets occurring after the mean response latency. Thus, there is a dramatic reduction in the degree of audiovisual interaction between the spatially congruent and incongruent conditions.

A.3.2.2 Congruent versus Incongruent Audiovisual ERPs

Considering the dramatic differences between the spatially congruent and incongruent differential ERPs, we then examined the effect of spatial incongruency on audiovisual integration. An alternative differential ERP was calculated, comparing the spatially congruent (A_iV_i or A_jV_j) and the spatially incongruent (A_iV_j or A_jV_i) ERPs:

$$ERP_t(A \otimes V) = [ERP_t(A_i, V_i) + ERP_t(A_j, V_j)] - [ERP_t(A_i, V_j) + ERP_t(A_j, V_i)]$$

where this time the $A \otimes V$ interaction represents the spatial effect. One possible confound with the first calculation of the differential ERP is that any increase in activity present for all stimulus conditions (A, V, or AV_X) would have been subtracted twice but only added once (AV_X - (V + A)) and could erroneously indicate an audiovisual interaction. With this formulation, any brain activity common to both audiovisual conditions—but not related to sensory interaction—would be eliminated; the confounding effect of unequal subtraction of a common component from the earlier formulation of audiovisual interaction can be avoided (Calvert & Thesen, 2004; Teder-Salejarvi, McDonald, Di Russo, & Hillyard, 2002). The temporal profile of statistically significant audiovisual interaction was calculated, as earlier, by running t-tests.

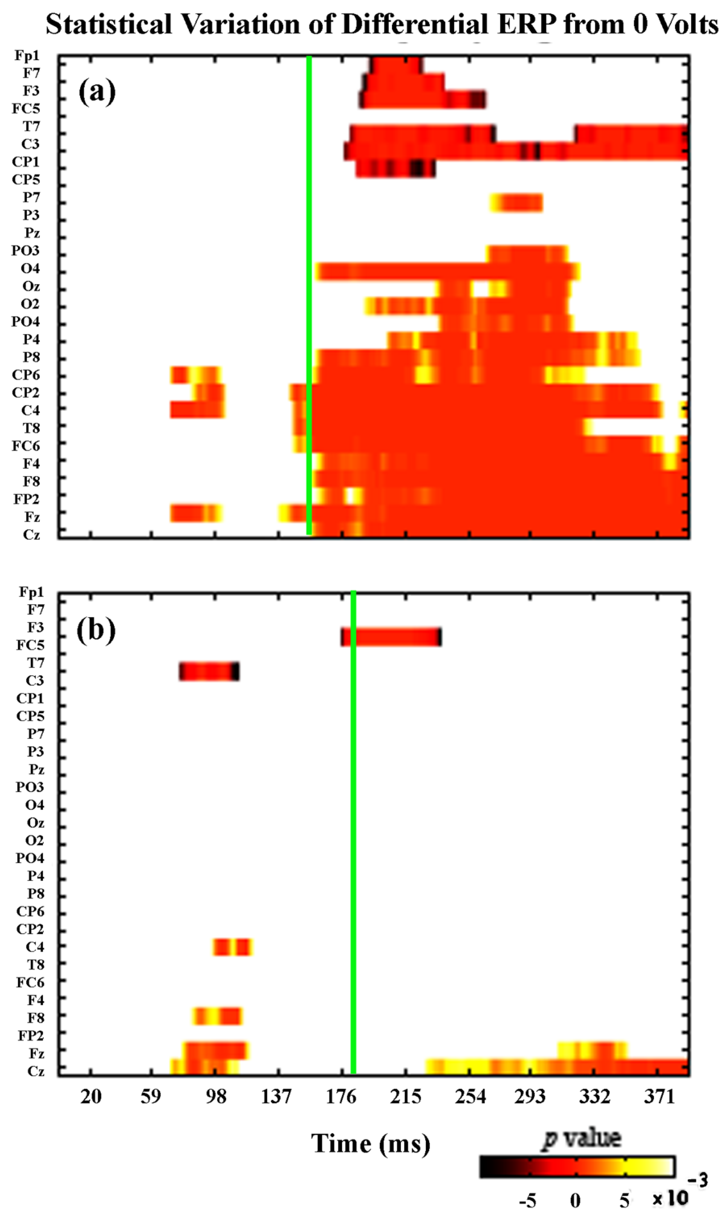

Figure A.2: Plots of differential ERPs ($ERP(AV) - [ERP(A) + ERP(V)]$) for spatially (a) congruent and (b) incongruent audiovisual stimuli that were significantly different from 0 volts at each of the 28 electrode positions, post-target onset. Solid green line indicates time of mean response latency found in behavioral data.

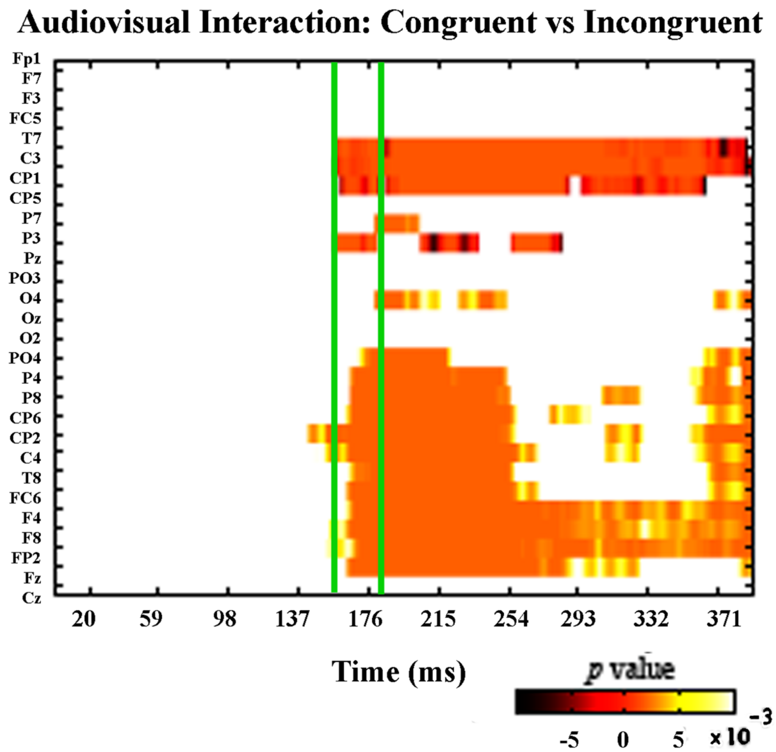

Figure A.3: Plot of difference between spatially congruent and incongruent differential ERPs ($ERP(AV_C) - ERP(AV_I)$). Green bars represent the time of the mean response latencies for the spatially congruent and incongruent conditions, respectively.

In a comparison between the ERPs for the spatially congruent and incongruent trials, the earliest disparity effects were found in the right centro-parietal region at 142 ms, followed by a robust and consistent interaction over many regions (except left frontal region). Significant interactions in the right posterior regions persisted until 250 ms post-onset whereas a continuous audiovisual interaction was found in the frontal, right fronto-central, and left temporo-central regions (Figure A.3).

A.3.3 Topography

Looking more closely at the topography of the audiovisual and congruency interactions, we re-plotted the absolute differential ERPs (bimodal response minus the sum of the unimodal responses) for both AV_C and AV_I for 30 ms time windows beginning at 60 ms post-onset (Figure A.4).

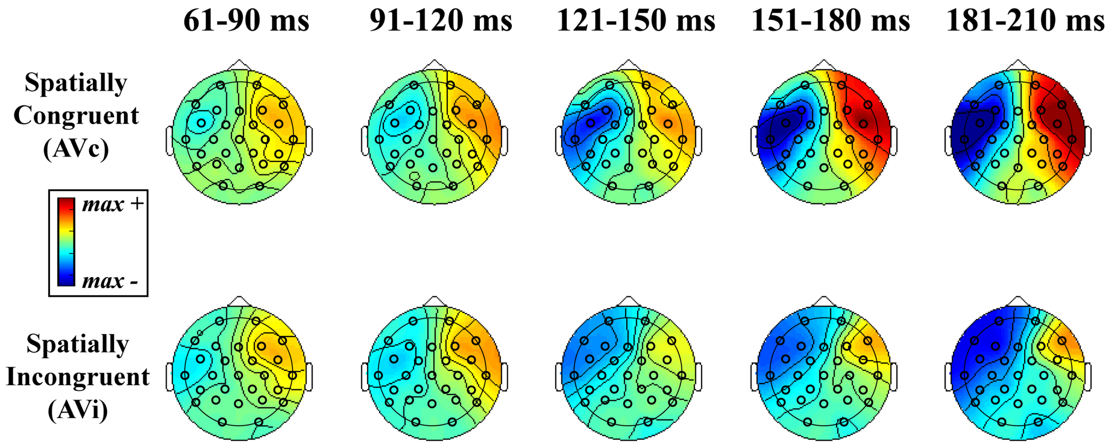

Figure A.4: Topographical plot of significant variance from 0 volts of spatially congruent and incongruent differential ERPs ($ERP(AV_X) - [ERP(A) + ERP(V)]$) in 30 ms time windows post-target onset.

For both AV_C and AV_I conditions, gradually stronger enhanced activations (supra-additive) were found in the right frontal and temporal regions from the time windows 61–120 ms, but after this time period the two diverge. Under AV_C conditions, this area of enhanced activation intensifies over the next 90 ms and is joined by an increasingly stronger suppression (sub-additive) in contralateral regions. Conversely, for AV_I conditions the activation in the right frontal and temporal regions did not continue beyond 120 ms and while the suppression in the left hemisphere—also beginning around the same time as in the spatially congruent condition—persisted, with a more diffuse area of suppression throughout the left frontal and central regions.

A.4 Discussion

In this study, we examined the effects of spatially congruent and incongruent audiovisual stimuli, compared to their unimodal constituent parts, in healthy adults using both behavioral and elecroencephalographic recordings. There was a consistency of results between both, revealing significant, intense and broad supra-additive activations in the right fronto-temporal regions coupled with suppression in the contralateral regions, and culminating in a faster mean response latency toward spatially congruent audiovisual stimuli at 25°. Spatially incongruent stimuli also differ from what would be expected by a simple additive effect between the unimodal responses, with a similar early rise in the right fronto-temporal area (91–120 ms), that quickly diverges from the congruent condition profile to non-significance, though with a broad and more weakly suppressive effect in the left frontal and temporal areas, resulting in a mean response latency that is comparable to latencies found for unimodal targets at the same location.

This adult study, while useful in it's own right for showing direct neural correlates and timing information and linking them to known overt response behavior in an audiovisual spatial localization task, also provides a sound basis for similar elecroencephalographic recordings in the infant brain. Given the handicaps that current developmental studies face in exploring multisensory integration in infants (inability to give instructions and lack of direct communication), EEG provides a way of making direct measurements on how an infant's brain responds to bimodal and unimodal stimuli.

Previous developmental studies have found evidence that the localization response in very young infants (under 5 months) is subcortically mediated, by the lack at this age of the "precedence effect"—perception of differential arrival times between bilaterally presented auditory stimuli—that neurophysical studies have shown to be cortically controlled (Clifton, Morrongiello, Kulig, & Dowd, 1981a; Clifton et al., 1981b). This, combined with our own behavioral data showing large developmental changes over the first ten months—such as the onset of limited, non-linear audiovisual integration at 8–10 months of age, or the changes in the inhibitory effects of spatial and temporal disparity between four and eight months—suggest profound neurological changes are also taking place during this time.