Investigating the validity of the additive model as a control in audiovisual integration studies

---

A Thesis

Presented to

The Division of Philosophy, Religion, Psychology and Linguistics

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Orestis Papaioannou

May 2015

Approved for the Division

(Psychology)

_____

Enriqueta Canseco-Gonzalez

Acknowledgments

There have been many people that have been instrumental in my life as a student and a senior. Too many to fit in this space. But I'd like to acknowledge the most prominent people that come to mind. My parents and family, for everything they sacrificed to bring me here. Chris for initiating me into the world of EEG, and for the countless hours spent talking science. Enriqueta and Michael for being wonderful mentors, and the rest of the SCALP lab for their help and company. Hayley for bringing me snackies while I worked, and for keeping me sane in these times of insanity. Ami for being there when it mattered, and Ian for sparking my academic mind and for being an all-around awesome person. And all the people that remain unmentioned, for you have made me who I am, and I would not be here without you.

Table of Contents

Introduction: 1

Audiovisual integration 1

Neuroimaging studies 1

Electrophysiological studies 2

The additive model 2

ERP results emerging from the model 3

Criticism to the model 4

i. sA and sV are most likely not independent 4

ii. There are often differences in task and attention demands between conditions 5

iii. There might be audiovisual interactions that do not reflect true audiovisual integration (i.e. interaction is **not** equal to integration) 6

Responses to Criticism 6

Investigating the validity of the models 7

Methods 9

Participants 9

Stimuli 9

Procedure: 13

ERP Recording: 13

Results 15

Behavioral Results 15

ERP Results: 16

Discussion: 19

Appendix A: Electrode Locations 23

Bibliography 25

List of Figures

Figure 1: Audiovisual interactions reported in previous studies. 3

Figure 2: The 4 different types of stimuli (A,V,AVc, and AVi), and their constituent unimodal parts. 10

Figure 3: Visual, auditory, audiovisual congruent and audiovisual incongruent stimuli . 11

Figure 4: Behavioral results. 15

Figure 5: ERP results. 16

Figure 6: Mass Univariate Analysis 17

Figure 7: Locations of the electrodes used. 23

Abstract

The additive model, where one compares the neural signal elicited by bimodal stimuli to the sum of the neural signals elicited by the independent presentation of the constituent unimodal stimuli, has been widely used in electrophysiological studies of audiovisual integration. However, the validity of the additive model is questionable, and has yet to be experimentally investigated. We compared the signals elicited by both congruent (integrated) and incongruent (and presumably not integrated) audiovisual stimuli to the summation of the signals elicited by unimodal auditory and visual stimuli to test if the additive model produced similar results to a direct comparison of the incongruent and congruent audiovisual stimuli. We found a posterior positivity at 75ms from stimulus onset when comparing audiovisual stimuli to the sum of the unimodal stimuli, and a later (~300ms) frontal positivity and posterior negativity when comparing the congruent audiovisual stimuli to either the incongruent audiovisual stimuli or the sum of the unimodal stimuli. We believe that the early effect reflects non-integratory processes occurring during simultaneous processing of two modalities, while the later effect reflects true audiovisual integration processes. However, the lack of differentiation in behavioral measures between the two types of audiovisual integration leaves the interpretation of our results open for discussion.

To everything that brought me here, and to wherever I go next.

Introduction:

Audiovisual integration

Audiovisual integration refers to the non-parallel simultaneous processing of auditory and visual stimuli, resulting in a single unified percept (e.g., perceiving the image of a face and the sound of a voice as an integrated stimulus of a face speaking, rather than as two separate stimuli). By definition, audiovisual integration assumes cross-communication between the auditory and visual streams, and seems to differ, both from unimodal auditory and visual processing, and from the simple union of the two. Evidence for this comes from many areas of psychology, using a variety of techniques.

One of the first pieces of evidence of audiovisual integration was the so-called McGurk Effect (McGurk & McDonald, 1974), which demonstrated that visual cues in speech affect our auditory percept and can even produce an illusory percept (e.g. hearing a person producing an auditory /ba/ while seeing an image of that person producing a visual /ga/, often induces an illusory percept of /da/). Similarly, Shams et al. (2000) showed that presenting a single flash with two rapid beeps would induce an illusory second flash temporally matching the second beep. Audiovisual integration also seems to have a facilitatory effect on stimulus processing. Giard and Peronnet (1999), for example, report that reaction times in a discrimination task were significantly faster for audiovisual presentation of simple stimuli, as compared to unimodal auditory or visual presentations. This facilitation seems to imply cross-sensory interactions of some sort, as it cannot be adequately modeled by a parallel and independent processing model.

Neuroimaging studies

Audiovisual interactions have also been shown using various neuroimaging techniques. Functional magnetic resonance imaging (fMRI) studies, for example, have found that the superior temporal sulcus shows increased activation during audiovisual

tasks, compared to the combined activity of the two unimodal auditory and visual tasks (Degerman et al, 2006). In addition, transcranial magnetic stimulation (TMS) studies suggest a functional involvement of the posterior parietal cortex in the binding of the two modalities (Maniglia et al., 2012). Furthermore, single cell recording in animal models reveal cells in the superior colliculi that respond preferentially to multimodal stimuli (Meredith & Stein, 1986).

Electrophysiological studies

Evidence supporting audiovisual integration comes from electroencephalographic (EEG) studies as well. Event related potentials (ERPs) have been reported with a multitude of audiovisual effects, ranging from an early (40-90ms) right lateralized positivity (Giard & Peronnet, 1999), to a modulation of the N1 (Besle et al., 2009), to mid-latency (150-250ms) effects (Elmer et al., 2011), to later (>400ms) speech-specific effects (Baart et al. 2013). The broad time range of these effects suggest that audiovisual integration is an automatic, persistent process, affecting the early processing of the stimulus, but also contributing to later more complex functions.

The additive model

Most of the electrophysiological investigations of audiovisual integration make use of the additive model to assess the effects of integration above and beyond those of parallel processing of two separate modalities. This model states that, to find the effects of audiovisual integration, one simply needs to compare the signal elicited by audiovisual stimuli (sAV) to the sum of the signal elicited by unimodal auditory stimuli (sA) alone plus the signal elicited by unimodal visual stimuli (sV) alone. Given the additive properties of electrical signals, two simultaneous signals elicited from independent sources would simply add together to form a single signal distribution. Thus, assuming that sA and sV are independent, the theoretical signal received from audiovisual processing (sAV), without audiovisual integration, should look identical to the sum of sA and sV. Therefore, any differences found between sAV and (sA+sV), whether reflecting

subadditivity (sAV<(sA+sV)) or superadditivity (sAV>(sA+sV)), can be interpreted as a result of audiovisual integration.

ERP results emerging from the model

The additive model has been used by different labs, with different analysis methods, and under different tasks. Given how variable the techniques and designs using the additive model are, comparing across studies is often problematic, with different patterns of results emerging.

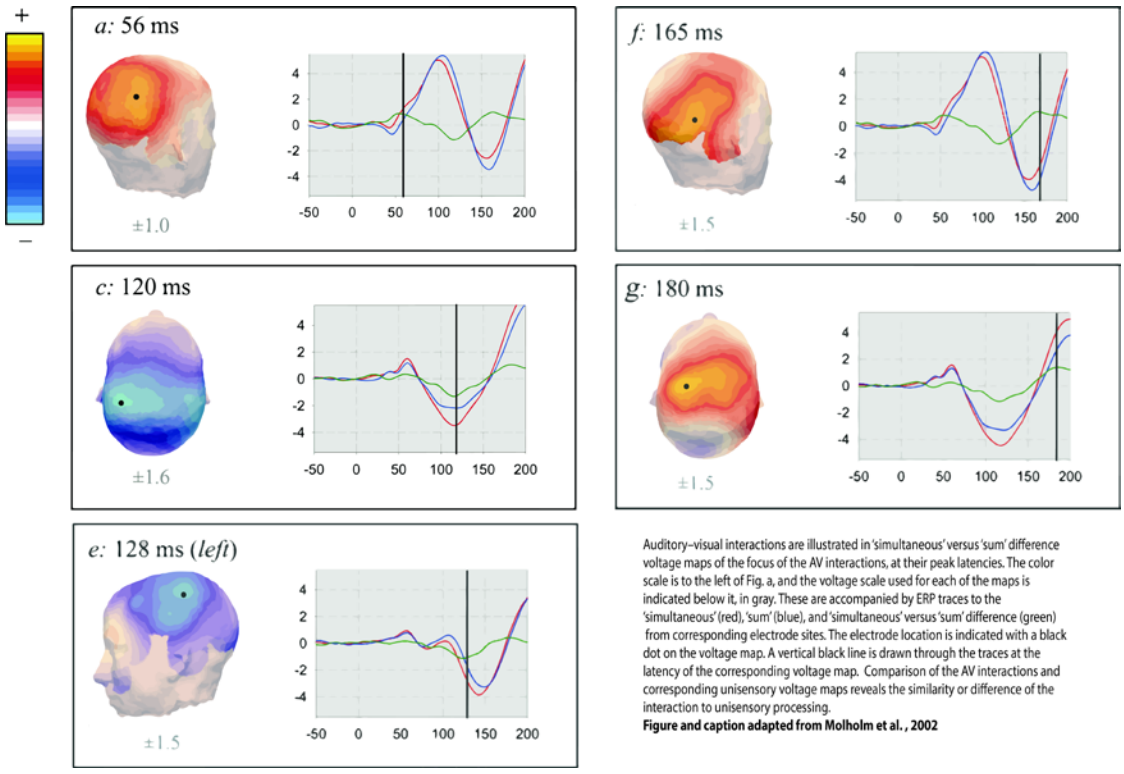

Figure 1: Audiovisual interactions reported in previous studies.
Difference maps and ERPs from Molholm et al. 2002, comparing the signal from audiovisual trials to the sum of the signals of the unimodal auditory and visual stimuli. Panels b and d were removed from the original because they were deemed irrelevant.

However, in general, two main time periods of interest emerge: a very early posterior positivity (40ms-100ms), and various early/mid latency effects (120ms-200ms),

with the former being the more robust of the two (see Fig 1). The very early effect manifests as a posterior positivity peaking at 60-70ms, and it was found using C1 as the reference electrode, with stimuli that were attended but not targets (Cappe et al., 2010), as well as with a nose electrode as the reference, with stimuli that were targets in a detection (Molholm et al., 2002) or discrimination (Giard et al., 1999) task. The early/mid latency effects are more variable in nature, but generally feature a central negativity at 120ms that becomes progressively lateralized, and reverts to a central positivity between 155ms-200ms (Molholm et al., 2002, and Giard et al., 1999). Importantly, these effects were not found in the Cappe et al. paper, where the AV stimuli were not targets (but were task relevant).

Criticism to the model

While widely used, the additive model is not without its faults. Given the complexity of the signal generators and the effects of higher cognitive processes across modalities, a few issues arise. The main three issues to consider are:

i. sA and sV are most likely not independent

While being primarily driven by auditory and visual processing respectively, sA and sV almost certainly contain signals originating in common circuits, such as attentional modulation, working memory circuits, motor responses (depending on the task), or any higher cognitive process that would be common to both types of stimuli. Since these effects are present in both signals, adding sA and sV together would effectively double these common processes.

Imagine, for example that there is an ERP component that reflects general conscious perception (GCP). Then, our stimuli would produce a signal that reflects both the processing of each respective stimulus and the corresponding GCP component, resulting in sA+GCP for auditory stimuli, sV+GCP for visual stimuli, and sAV+GCP for audiovisual stimuli. Then, subtracting the sum of the signal elicited by the auditory and visual stimuli, from that elicited by the audiovisual stimuli, we get (sAV+GCP)-[(sA+GCP)+sV+GCP)] = (sAV-(sA+sV))+GCP. Thus,

our difference wave includes both any audiovisual interactions present (sAV-(sA+sV)), plus the GCP, and hence does not reflect audiovisual integration effects alone.

This would be a particularly problematic issue when investigating effects in the mid to late latencies (i.e. effects found >~200ms from stimulus onset), where most of these higher cognitive processes are likely to occur. However, even early on, some common processes might be present. For example, there might be some modality independent attentional effects that are present very early after stimulus onset, which could possibly introduce early artifactual differences in the sAV-(sA+sV) comparison. Thus, this problem cannot be dismissed, even in studies that look exclusively at early effects.

ii. There are often differences in task and attention demands between conditions

There is a multitude of top down effects that could affect the processing of stimuli, with attention and task being some of the more common. Unfortunately, it is very hard to equate attention and task across unimodal and bimodal stimuli. In a unimodal task, participants effectively have 3 different types of tasks, each possibly requiring different levels of attention. For example, during task relevant unimodal trials (e.g. responding to a sound when a sound is presented), the participants are performing the basic unimodal task, while in task irrelevant unimodal trials (e.g. responding to a sound when a light is presented) they are essentially asked to inhibit the stimulus information. In bimodal trials where only one modality is task relevant, they are performing the task in one modality, and potentially inhibiting the other, which becomes effectively a dual task. On the other hand, a task in both modalities would mean that bimodal trials could have a lower task difficulty, as the participant could rely on the most salient of the two modalities in each trial to perform the task (something that is not possible in unimodal trials).

Similarly, it is very hard to equate the attention paid to unimodal and bimodal stimuli, especially because audiovisual integration has a multifaceted relationship with attention. It has been shown, for example, that reaction times in a target

detection task were significantly faster for bimodal than unimodal targets, suggesting that audiovisual integration facilitates bottom-up attention modulation. However, studies also found that this effect was significantly reduced under a heavy attentional load, implying that a certain amount of top-down attention is needed for audiovisual integration to take place (for a comprehensive review, see Talsma et al., 2010).

*iii. There might be audiovisual interactions that do not reflect true audiovisual integration (i.e. interaction is **not** equal to integration)*

The additive model assumes that any differences between sAV and (sA+sV) reflect meaningful effects of audiovisual integration. However, it is possible that some of those differences are simple effects of simultaneous processing of each of the two modalities and not necessarily a result of integration. For example, the two modalities might be competing for a shared pool of resources (such as attention or working memory), and might not be processed at full efficiency in simultaneous multimodal presentations. Furthermore, in trials with unimodal stimuli, the participants may be actively trying to suppress the task irrelevant modality (suppressing the image of the fixation cross in auditory trials, the background noise in visual trials, etc). However, in bimodal trials, this suppression may either be absent completely, or change from complete suppression of a modality to only a partial one (suppressing only extraneous auditory and visual stimuli). It is hard to imagine this kind of differential processing being part of the definition of audiovisual integration, as it could occur with any combination of stimuli and do not necessarily result in a binding of the two streams, but rather in general, non-integratory interactions.

Responses to Criticism

These issues have been a topic of discussion in the literature, and have yet to be resolved. Besle et al. (2009), for example, discuss these problems and offer ways to minimize their effects. Specifically, they mention that researchers using this model should: a) focus on the first 200ms of the signal to minimize the effects mentioned above

in #1, since most common processes are thought to occur later in time, and b) carefully manipulate the task and design to minimize the problems mentioned in #2, such as choosing a bimodal task that is equally difficult in each modality and avoiding blocked designs. While this is a step in the right direction, these guidelines are not universally accepted (Baart et al. 2013), they do not apply to research using techniques other than EEG (Degerman et al, 2006), and it is uncertain to what extent these measures alleviate the problems outlined above.

Investigating the validity of the models

While the additive model is widely used in studies of audiovisual integration, its validity has yet to be, to my knowledge, experimentally investigated. This study attempts to do just that. Our strategy was to use a new type of stimulus in addition to typical unimodal auditory (A) and visual (V) stimuli, and congruent audiovisual stimuli (AVc). We included incongruent audiovisual stimuli (AVi) that were unlikely to be integratable. That way, we were able to compare the additive model to an audiovisual control. Furthermore, a number of experimental features were controlled; the task was designed so that task relevance and task difficulty were equated across modalities, and the incongruent AV stimuli were made to resemble the congruent AV stimuli as much as possible.

By comparing AVi to AVc, the problems in #1 are circumvented, since the comparison does not assume complete independence between modalities. Similarly, any general non-integratory AV interactions should occur in both the integratable and non-integratable stimuli, nullifying the concerns raised in #3. Furthermore, by choosing a task that meets the above criteria, the task-related effects should be minimal, while the similarity between the congruent and incongruent stimuli should ensure that non-integrative higher processes (namely attention and working memory) do not differ substantially between conditions.

If the additive model is valid and approximates simultaneous non-integrated processing, then comparing AVc to the addition of the unimodal stimuli (A+V), should produce a signal difference that is very similar to that obtained when comparing AVc to

the signal produced by the non-integratable audiovisual stimuli (AVi). However, if some of the effects found using the additive model are missing when comparing AVc to AVi, then those effects are likely a result of extraneous artifacts produced by the above mentioned models.

Methods

Participants

14 Reed College students (7 female, mean age = 21 years old) with no history of neurological trauma participated in the study. Three more participants (2 female) were excluded due to failure to perform above chance at the task, or to excessive EEG artifacts. The participants were awarded departmental lottery tickets for their participation, for a chance to win $50. Informed written consent was collected from all participants, and the experimental procedures were approved by the Reed College institutional review board, in compliance with the Declaration of Helsinki.

Stimuli

Our goal was to create two types of AV stimuli: one in which the auditory and visual stimuli were easily integrated, and one where (although also simultaneously presented), they were not perceived as integrated. Furthermore, unimodal stimuli were also created, consisting of either the auditory and visual components of the AV stimuli (see fig 2).

Six auditory and six visual stimuli were used. The auditory stimuli consisted of two categories: three 50ms tones (of 360Hz, 440Hz, and 520Hz respectively) presented at 74dB (from here on ‘Beeping Tones’; see fig 3.C), and three 500ms tones (of the same three frequencies as above) with an ascending volume starting at 68dB and ramping up to a maximum of 74dB at 450ms (from here on ‘Sliding Tones’; see Figure 3.D).

		Auditory (A)	Visual (V)	Audiovisual Congruent (AVc)	Audiovisual Incongruent (AVi)
Stimulus A	Visual Stimulus	-	Flashing Dots	Flashing Dots	Sliding Dots
	Auditory Stimulus	Sliding Tone	-	Beeping Tones	Beeping Tones
Stimulus B	Visual Stimulus	-	Sliding Dots	Sliding Dots	Flashing Dots
	Auditory Stimulus	Beeping Tones	-	Sliding Tone	Sliding Tone

Figure 2: The 4 different types of stimuli (A,V,AVc, and AVi), and their constituent unimodal parts.

The tones from the first category were grouped into three possible presentation sequences:

a) A 360hz (low) tone, followed by a 520Hz (high) tone, and then a 440Hz (mid) tone
b) A 440hz tone (mid), followed by a 520Hz (high) tone, and then another 440Hz (mid) tone
c) A 520Hz (high) tone, followed by another 520Hz (high) tone, and ending with a 360Hz (low) tone

In all three sequences, the stimuli where presented at 0ms, 210ms and 375ms respectively.

Visual stimuli were similarly separated into 2 categories of three trial types each. The first group consisted of three white circles, each 50ms in duration, and 0.65° in diameter, presented at -1.5°, 0°, and 1.5° from the center along the vertical axis (from here on 'Flashing Dots'; see Figure 3.A). These stimuli were grouped in similar sequences as the 50ms tones, with the 360Hz, 440Hz, and 530Hz tones replaced by the -1.5°, 0°, and 1.5° circles respectively. The second group consisted of two white circles, 0.49° in diameter, moving away from each other either horizontally, diagonally, or vertically (from here on 'Sliding Dots'; see Figure 3.B). The circles moved at a steady rate for 500ms, starting at 0.25° separation and ending 5.8° from each other.

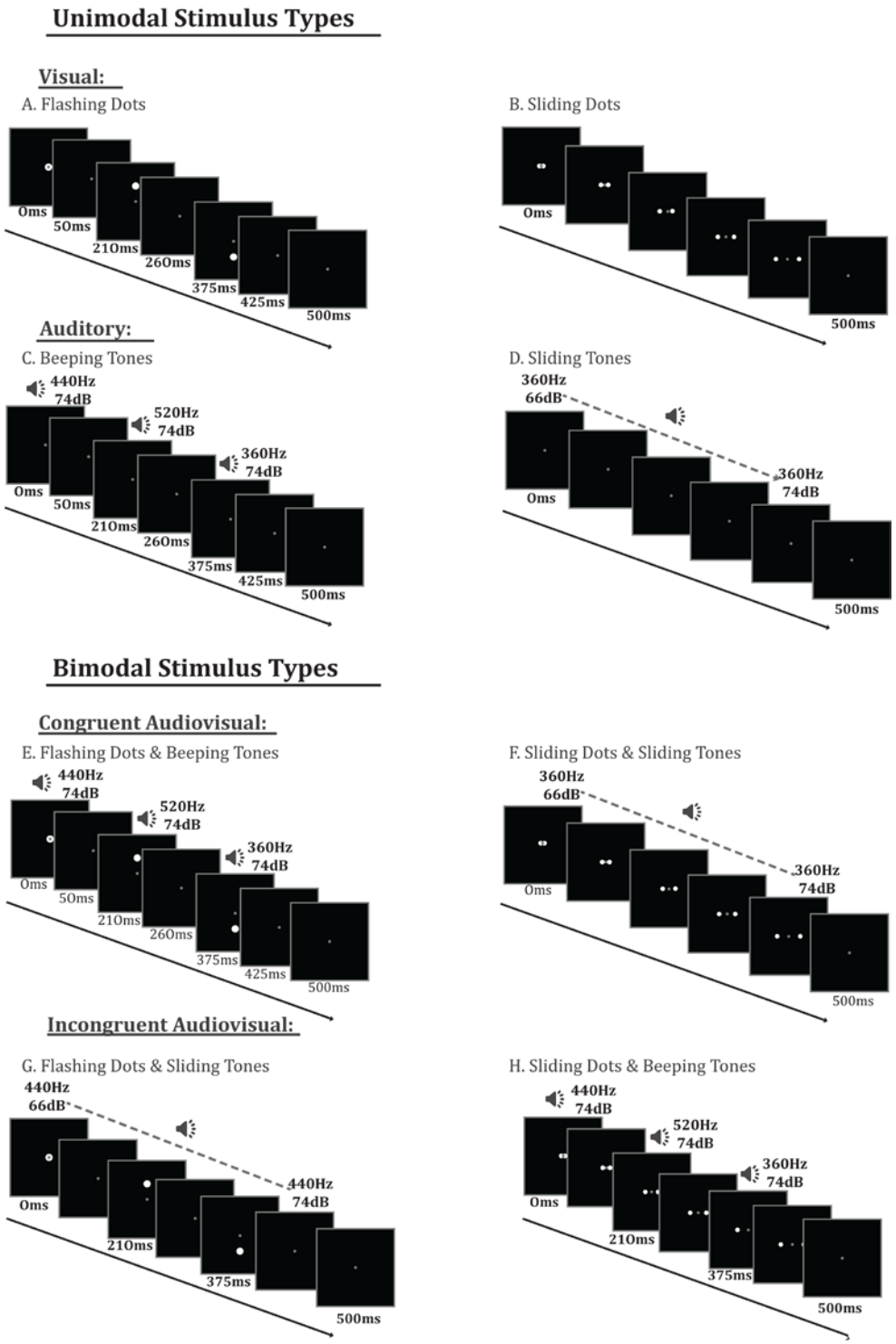

Figure 3: Visual, auditory, audiovisual congruent and audiovisual incongruent stimuli
A temporal representation of the different stimuli is shown. The Sliding Dots moved continuously away from the center of the screen, while the Flashing Dots remained on screen for 50ms at a time. Similarly, the Beeping Tones were presented for 50ms each, but the Sliding Tones were presented for 500ms.

These stimuli were combined to form 3 types of trials: unimodal trials, audiovisual congruent, and audiovisual incongruent. The first consisted of either the auditory or visual stimuli presented unimodally. The second consisted of congruent stimuli; Beeping Tones being presented simultaneously with the Flashing Dots (e.g. 3 tones perceived as “high-low-medium” paired with 3 circles appearing at high-low-medium vertical locations), or the Sliding Tones being presented simultaneously with the Sliding Dots (e.g. a 360Hz, 500 ms tone increasing in volume, paired with the image of two dots moving away from each other horizontally over 500ms) making an audiovisual stimulus that is readily integrated (i.e. elicits a subjective experience of the auditory and visual stimuli as being a congruent stimulus) (see fig. 3.E and 3.F). The last type consisted of incongruent stimuli; Beeping Tones and Sliding Dots, or Sliding Tones and Flashing Dots, presented simultaneously, producing audiovisual stimuli that were incongruent (for example, 3 tones of different pitch, paired with an image of two dots moving away from each other) (see fig. 3.G and 3.H).

A pilot study with 4 participants was performed to verify that they perceived the audiovisual congruent stimuli (AVc) as unified percepts, but were unable to do so for the audiovisual incongruent (AVi) stimuli. In all audiovisual combinations, 360Hz tones were paired with the -1.5° circles or the horizontal moving 500ms circles, 440Hz tones were paired with 0° or diagonal moving circles, and 520Hz were paired with the 1.5° or vertical moving circles correspondingly. In cases where three 50ms tone or three 50ms circle sequences were paired with the 500ms stimulus of the opposing modality, the first stimulus in the sequence matched the 500ms stimulus in the way mentioned above

Additional target stimuli, corresponding to the 3 types of experimental stimuli mentioned above, were created by reducing the brightness of the visual stimuli and the volume of the auditory stimuli by 25% and 6dB respectively. Pilot studies confirmed that the targets were equally hard to detect in both modalities. Finally, a fixation dot of a 0.1° diameter was centrally presented throughout the experimental session. All stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, CA).

Procedure:

The experiment consisted of 14 identical two-minute blocks, each consisting of 120 non-target trials. Each of the 4 trial types (lasting 500ms each) was presented a total of 4 times per block, in a randomized order, with a 400ms-600ms interval between trials. Additionally, target trials were randomly interspersed between non-target trials, adding up to 10% of all presentations. Participants were asked to press a button whenever they detected one of the target trials. A short break was given to the participants at the end of each block.

ERP Recording:

Brain electrical activity was recorded at the scalp using a Herrsching DE-82211 "Easycap" with 28 electrodes placed using the standard 10-20 system (for a detailed schematic of electrode position, see Appendix A). Additional electrodes were placed on the left and right mastoids, on the outer side of the left and right eye, and below the left eye, as well as a ground electrode on the FC4 position. A high sodium gel was used to establish a conductive bridge between each electrode and the scalp, and electrode impedances were kept below 5kΩ. Electrode signals were sampled at a 500Hz digitization rate and amplified by BrainVision "Professional Brain-Amp" amplifiers. A low pass filter of 150Hz, 24db/octave and a high pass 0.1Hz, 24db/octave filters were applied online. A 30Hz, 24db/octave low pass filter was applied to the data during offline processing. The electrodes were referenced online to the right mastoid, and re-referenced offline to the average of the 28 electrodes on the cap. Trials with blinks or eye-movements were identified and removed using the data from the additional electrodes placed near the eyes. Trials with excessive noise in other electrodes were also removed. The average number of trials remaining after artifact rejections was 340 for A, 362 for V, 360 for AVi, and 351 for AVc.

Results

Behavioral Results

Our behavioral results can be seen in Fig. #4 below. There were no significant differences in reaction time (M = 660.3 and 657.5 respectively, $t(13)= 0.331$, $p>0.1$) or accuracy (M = 95.2% and 95.3%, $t(13)= -0.161$, $p>0.1$) for AVc and AVi respectively.

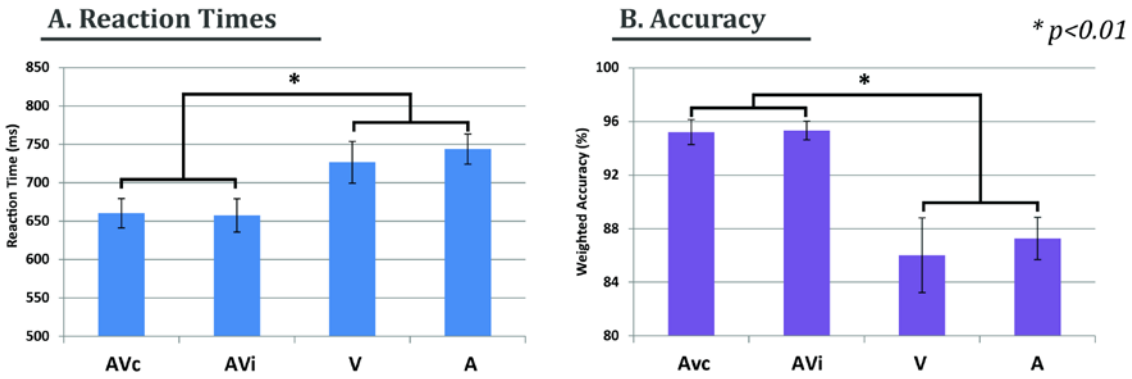

Figure 4: Behavioral results.
Mean reaction time (A) and accuracy (B) for each stimulus type. Accuracy was calculated by taking the average between (correct rejections/false alarms) and (correct hits / misses). Chance is at 50%.

Similarly, performance for visual (mean RT = 726.7, mean accuracy = 86.0%) and auditory (mean RT = 743.8, mean Accuracy = 87.2) unimodal stimuli did not differ from each other either in reaction time ($t(13) = -1.11$, $p>0.1$) or accuracy ($t(13) = 0.122$, $p>0.1$). However, unimodal visual stimuli produced significantly slower reaction times and lower accuracy than AVc ($t(13)= -2.99$, $p<0.01$, $t(13)= 3.29$,$p<0.01$) and AVi ($t(13)= -3.02$, $p<0.01$, $t(13)= 3.27$,$p<0.01$). The same was true for unimodal auditory stimuli ($t(13)= -8.21$, $p<0.01$ and $t(13)= 5.54$, $p<0.01$ for AVc, and $t(12)= -6.94$, $p<0.01$ and $t(13)= 6.66$,$p<0.01$ for AVi). To summarize, no significant differences in performance

were found between the two types of AV stimuli, or between the two unimodal stimuli, but performance was overall faster and more accurate for audiovisual stimuli.

ERP Results:

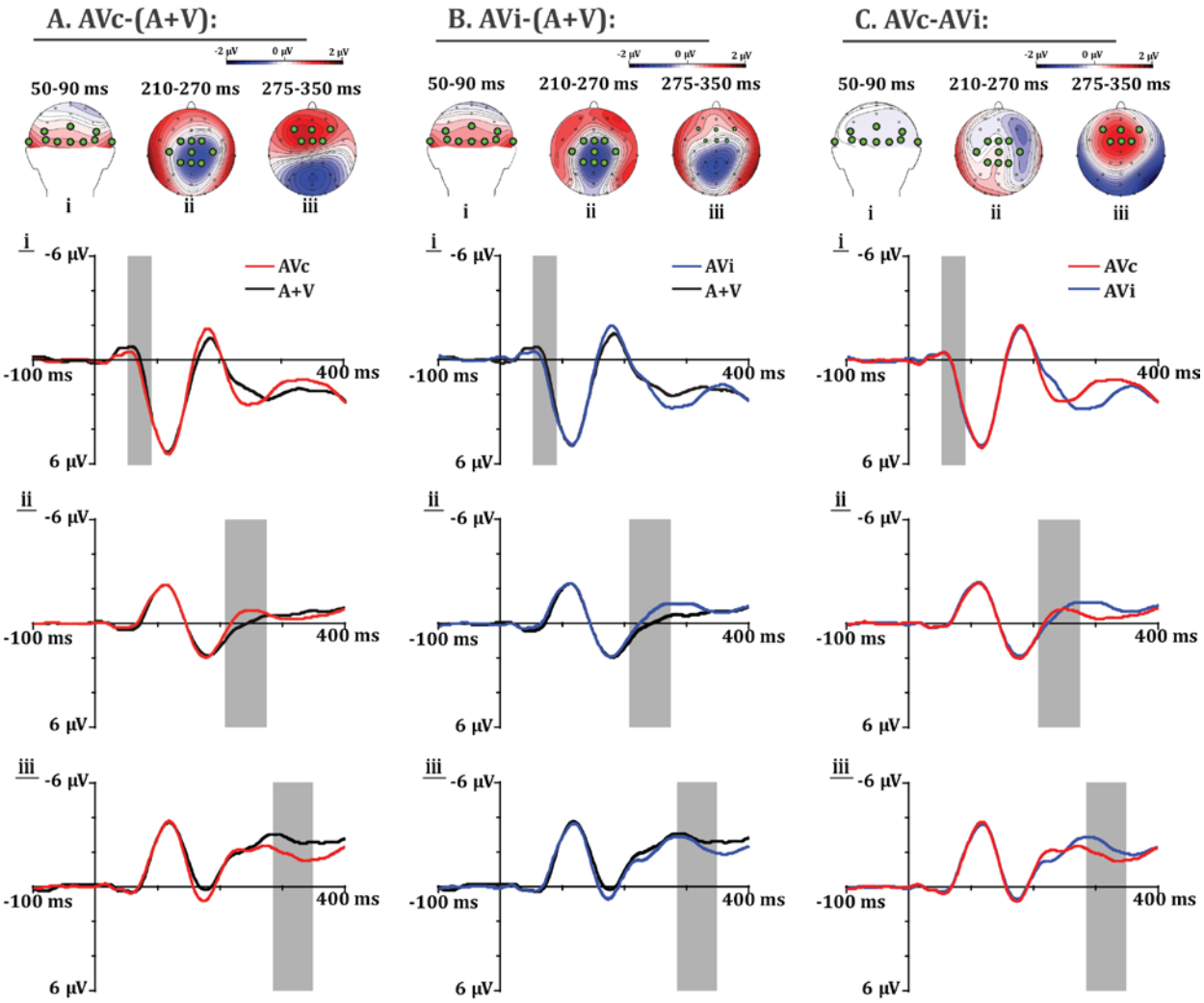

Figure 5: ERP results.

ERPs and difference maps for our different comparisons. Three time windows of interest are highlighted: 50-90ms (i), 210-270ms (ii), and 275-350ms (iii). The highlighted electrodes in the difference map of each time window were pooled together to create the ERPs shown.

In order to assess the additive model of sensory integration, we first calculated three separate difference waves: AVc-(A+V), AVi-(A+V) and AVc-AVi (see fig. 5). We then submitted these difference waves to a Mass Univariate analysis allowing us to test multiple time points and electrodes while correcting for multiple comparisons. Our first analysis was a cluster permutation test for all electrodes in the time window of 0ms-200ms which revealed that there were no significant effects in this time window.

However, a bilateral posterior positivity approached significance in both AVc-(A+V) and AVi-(A+V) at the mastoid electrodes peaking at 75ms ($p<0.08$) (see A.i and B.i in fig. 5). If we were to use a more traditional t-test to analyze this time period, then this posterior positivity becomes significant for AVc-(A+V) ($t(13)= 2.4$, $p<0.05$) and AVi-(A+V) ($t(13)= 2.9$, $p<0.05$), but not for AVc-AVi ($t(13)= -0.69$, $p>0.1$).

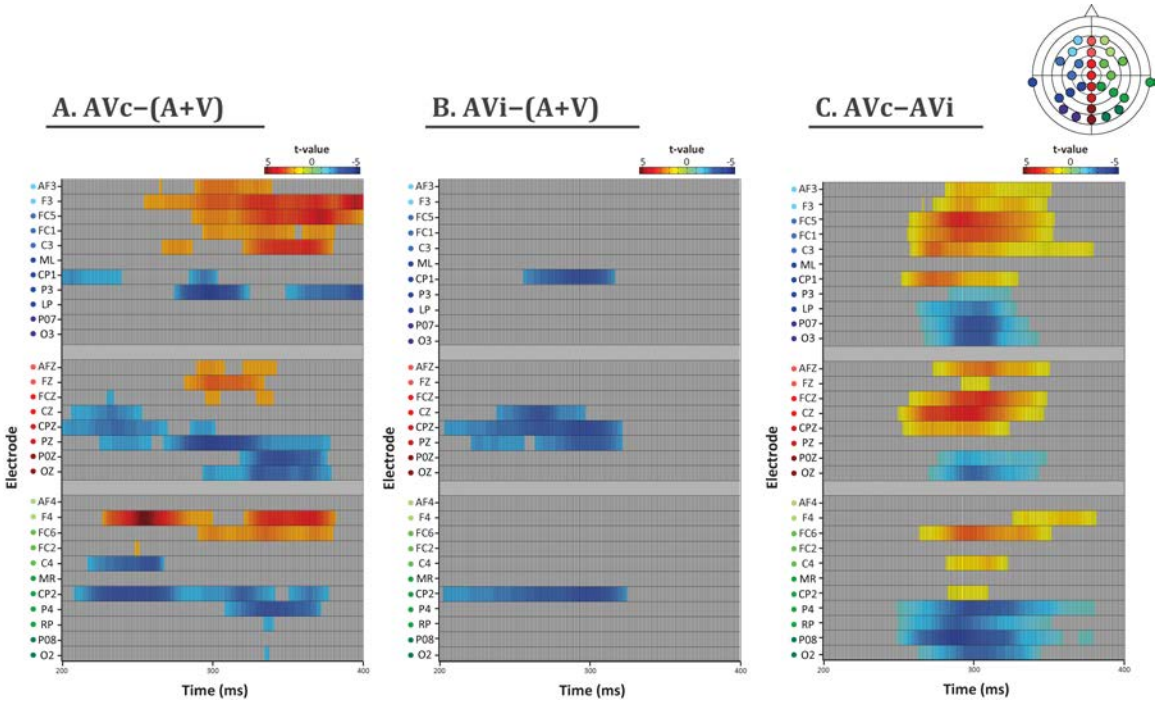

Figure 6: Mass Univariate Analysis

Results of the cluster permutation test for three comparisons. Each column represents a time point, and each row an electrode. Colored cells indicate significant differences.

While most studies using the additive model do not analyze any data past 200ms, due to the fact that the validity of the model decreases substantially at later time windows, our AVc-AVi comparison does not have such limitations. Thus, a cluster

permutation test for the same difference waves and for all electrodes was performed for the 200-400ms time window, allowing us to directly compare AVc-AVi to AVc-(A+V) and AVi-(A+V). (see fig. 5 and fig.6). This analysis revealed a significant posterior negativity and a frontal positivity around 275-350ms after stimulus onset for AVc-(A+V) and AVc-AVi only (see A.iii and C.iii in Fig. 5). Furthermore, a significant central negativity was found between 210-280ms in AVc-(A+V) and AVi-(A+V), but not for AVc-AVi (see A.ii and B.ii in Fig. 5).

Discussion:

Our behavioral results replicate previous findings of facilitated processing of audiovisual stimuli, with participants responding quicker and more accurately for audiovisual over unimodal stimuli. However, surprisingly and contrary to our expectations, we did not observe any significant differences between the two types of audiovisual stimuli, with participants having comparable reaction times and accuracies for both AVc and AVi. This complicates further interpretations of our results, since we cannot be certain that our stimulus manipulation worked as intended.

Recall that, in contrast with our congruent stimuli, we designed our incongruent stimuli such that they would not lead to sensory integration. However, it is possible that, after numerous repetitions of our incongruent stimuli, our participants learned to integrate AVi equally well to AVc, in which case the two would be equivalent in terms of audiovisual integration effects.

Alternatively, it could also be that under our specific task, the increased performance in audiovisual trials was solely due to the presence of two simultaneous information streams (visual and auditory), allowing the participants to rely on whichever modality is the most salient in any given trial. If this is the case, the actual integration of the two modalities did not significantly alter the behavioral performance on the task, meaning that it is still possible that our stimulus manipulation was indeed effective, with AVc being integrated while AVi was not (at least not to the same degree), implying that AVi is a suitable control for audiovisual integration, and thus differences between AVi and AVc reflect true audiovisual integration effects.

Given that significant differences were observed between AVc and AVi, the ERP results suggest that the two types of audiovisual stimuli are non-equivalent, and thus make the latter interpretation of our behavioral results (in which differences between AVc and AVi reflect real audiovisual integration effects) more likely. Assuming that this is true, some interesting conclusions arise from our data.

Firstly, the fact that neither we nor Cappe et al. (2010) observe any of the effects reported to occur between 120-200ms, may suggest that those effects are either observed only when using specific reference electrodes (with all the studies reporting those effects using a nose reference), or under discrimination or detection tasks that require a motor response, something that is absent in both Cappe et al.'s and our study. It is likely that these effects reflect some motor preparation processes that differ between AV and (A+V) in latency, due to the decreased reaction times exhibited in audiovisual trials.

The observed trending posterior positivity at 75ms (found in both AVc-(A+V) and AVi-(A+V)), on the other hand, could be interpreted as a replication of the very early effect that is commonly reported. However, the fact that no such positivity was found in the AVc-AVi comparison, but was consistently found when comparing audiovisual versus the sum of visual and auditory stimuli, would suggest that this is a result of simultaneous bimodal processing, but not of true integratory processes. While it is hard to know what the involved processes are, one possible explanation would be that these reflect attentional differences between stimulus types. For example, when unimodal stimuli are presented, the participant's full attention can be allocated to processing a single stream of information, which could lead to increased amplitude in the ERP markers of said processes. When processing audiovisual stimuli, however, the attention of the perceiver must necessarily be split between the two modalities, meaning that each individual modality is attended less, potentially leading to smaller amplitude ERPs compared to (A+V).

Previous studies utilizing the additive model generally do not analyze any differences past 200ms, since a lot of processes that are common to both A and V start showing effects at later time points. These common processes would lead to additional artifactual differences, and would make any results obtained very hard to interpret (Besle et al., 2009; Giard et al., 2010). However, the AVc-AVi comparison does not suffer any such limitations, which allowed us to make comparisons up to 400ms. AVc-(A+V) and AVi-(A+V) were also tested at that time interval, to allow for direct comparisons between the additive model and AVc-AVi. However, AVc-(A+V) and AVi-(A+V) cannot be used to make standalone conclusions.

An effect was found in AVc-AVi, manifesting as a posterior negativity and a frontal positivity around 300-400ms, which could be a true integration effect. A somewhat similar effect was found in AVc-(A+V) but not AVi-(A+V), further supporting the interpretation that this constitutes a true integration effect. This would place audiovisual integration effects much later than previous literature would suggest, and would change the perception of audiovisual integration from an automatic, early sensory effect to one happening after basic sensory processing. Some later effects were also found in AVc-(A+V) and AVi-(A+V), but, for the reasons stated above, these are just as likely to be artifactual rather than true effects, and thus we will not focus on them.

It is important to note, however, these interpretations rely heavily on the assumption that AVi were not integratable, which is not necessarily true. While the differences between AVc and AVi suggest that the two stimulus types are not equivalent, their differences may be based on some other property beyond integratability (such as the similarity between the temporal signatures of the two modalities) and thus would bear little relevance to the current study. If that is the case, then any findings in AVc-AVi are not due to audiovisual integration, but then those found in AVc-(A+V) and in AVi-(A+V) could be. This would mean that the early posterior positivity (found when comparing audiovisual stimuli to the additive model) may possibly reflect true audiovisual integration effects. It is equally likely, nonetheless, that they reflect similar non-integratory cross modal interaction effects to those stated in the initial interpretation, with the current results not being able to differentiate between the two.

In sum, while the exact interpretation of our results remains inconclusive, this study, at the very least, calls for further investigation of the additive model. We failed to replicate various results that were previously reported as audiovisual integration effects, raising suspicions that such effects could be a result of either, experimental design, or post-recording analysis. Even the effects that were replicated are not necessarily audiovisual integration effects, and could instead be an artifact of simultaneous processing of multiple modalities, something that the additive model cannot adequately distinguish. Lastly, it seems that the limited window in which one can look for effects using the additive model may be too limiting, hiding possible differences that occur post 200ms, such as those observed in this study.

Further research could answer some of the issues raised above, but it will require careful stimulus control to ensure simultaneously presented audiovisual stimuli that remain non-integrated even after hundreds of presentations. These may be achieved by using either stimuli with very contrasting characteristics between modalities, or perhaps using more complex stimuli with very strong audiovisual associations (e.g. a face with speech, compared to a saxophone with music). Thus, the validity of the additive model could be determined more conclusively, allowing electrophysiological investigations of audiovisual integration to gather more accurate results, and furthering the understanding of these integratory processes.

Appendix A: Electrode Locations

The locations of the electrodes used are shown below. We used 26 electrodes in the standard 10-20 system (plus one more for our ground), and 4 electrodes that were in non-standard locations. These were on the left and right mastoids (ML and MR respectively) and the left and right parietal electrodes (LP and RP respectively). In addition, 3 EOG electrodes were used (not shown in the figure).

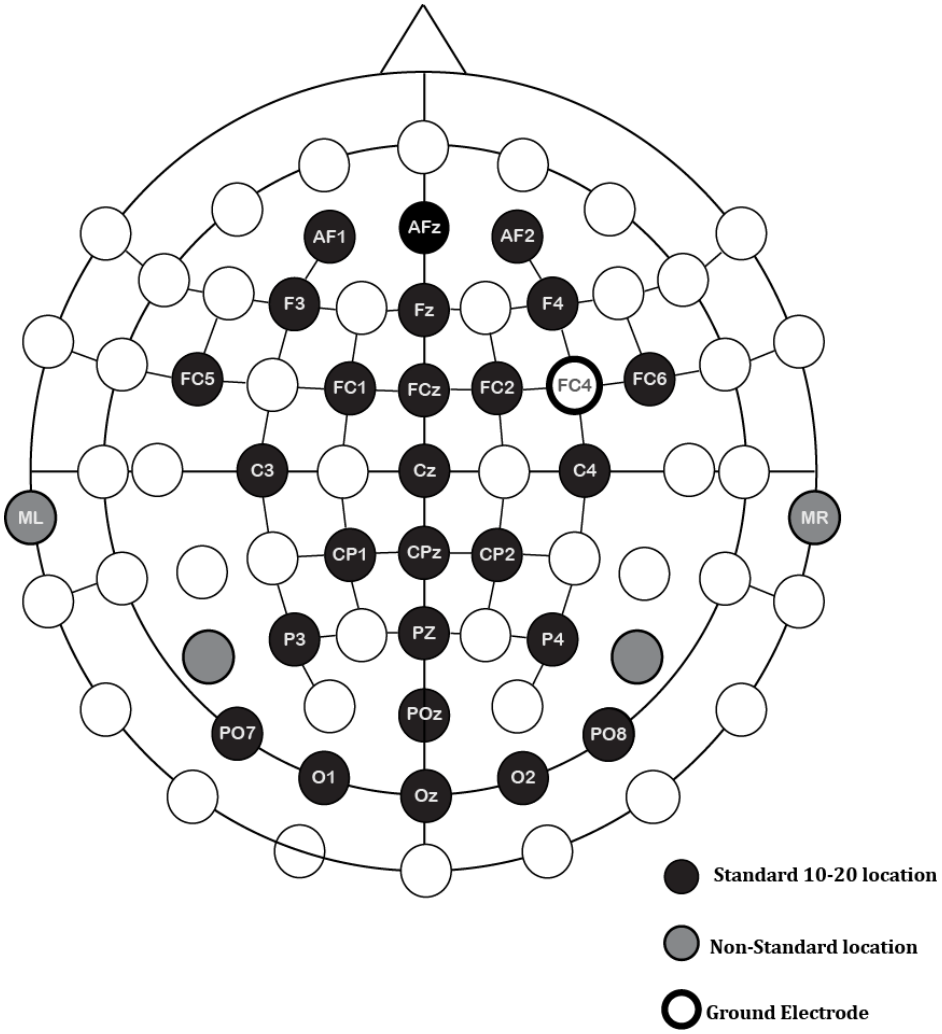

Figure 7: Locations of the electrodes used.

Bibliography

Baart, M., Stekelenburg, J.J., Vroomen, J., 2014. Electrophysiological evidence for speech-specific audiovisual integration, Neuropsychologia, Volume 53, 115-121.

Besle, J., Bertrand, O., Giard, MH., 2009. Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex, Hearing Research, Volume 258, Issues 1–2, 143-151.

Cappe, C., Thut, G., Vincenzo, R., Murray, M.M., 2010. Auditory–Visual Multisensory Interactions in Humans: Timing, Topography, Directionality, and Sources, The Journal of Neuroscience, Volume 30, 12572-12580.

Degerman, A., Rinne, T., Pekkola, J., Autti, T., Jääskeläinen, I.P., Sams, M., Alho, K., 2007. Human brain activity associated with audiovisual perception and attention. NeuroImage, Volume 34, Issue 4, 1683-1691.

Elmer, S., Meyer, M., Jäncke, L., 2012. The spatiotemporal characteristics of elementary audiovisual speech and music processing in musically untrained subjects, International Journal of Psychophysiology, Volume 83, Issue 3, 259-268.

Giard, M.H., Besle, J., 2010. Methodological Considerations: Electrophysiology of Multisensory Interactions in Humans. Multisensory object perception in the primate brain, Naumer (Ed.), 55-70.

Giard, M.H., Peronnet, F., 1999. Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. Journal of Cognitive Neuroscience. 11, 473–490

Maniglia, M., Grassi, M., Casco, C., Campana, G., 2012. The origin of the audiovisual bounce inducing effect: A TMS study, Neuropsychologia, Volume 50, Issue 7, 1478-1482.

McGurk, H., McDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746– 748.

Meredith, M.A., Stein, B.E., 1986. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. Journal of Neurophysiology 56, 640–662.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cognitive Brain Research, Volume 14, 115-128.

Shams, L., Kamitani, Y., Shimojo, S., 2000. Illusions. What you see is what you hear. Nature, 408, p. 788.

Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multisensory integration, Trends in Cognitive Sciences, Volume 14, Issue 9, 400-410.