UNIVERSITY OF CALIFORNIA, SAN DIEGO

Neural Processes underlying an Auditory-induced Visual Illusion

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology/ Specialization Computational Neurobiology

by

Jyoti Mishra

Committee in charge:

Steven Hillyard, Chair
Karen Dobkins
Eric Halgren
John Reynolds
Terrence Sejnowski

2008

The dissertation of Jyoti Mishra is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2008

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

ACKNOWLEDGEMENTS

This thesis could not have been accomplished without the exceptional guidance of my learned advisors, and the invaluable support of my inspiring colleagues, loving family and close friends. I would especially like to thank my thesis advisor, Dr. Steven A. Hillyard, for teaching me every aspect of becoming a good scientist. I am thankful to Dr. Terry Sejnowski and the Computational Neurobiology program for supporting me and mentoring me in computational methods. I also owe a special thanks to Antigona Martinez who has been a mentor and cherished friend at the same time. I express my gratitude to all members of my thesis committee for their invaluable guidance and thoughtful comments through the course of the thesis that have helped maintain the scientific rigor and quality of this work. I thank all members of the Hillyard lab for their enriching scientific interactions, and Matt Marlow and Carole Montejano for their respective technical and administrative support. Finally, and most importantly, I am very thankful to my family in India, who have always believed in my abilities and have inspired me to take on many challenges. I also thank Dhakshin Ramanathan for sharing all the exciting and difficult moments of scientific research with me, and am grateful to his loving family for giving me a home away from home.

Chapter 2, in full, is a reprint of the material as it appears in the Journal of Neuroscience, 2007, Mishra, J.; Martinez, A.; Sejnowski, T. J., Hillyard, S. A. The dissertation author was the primary investigator and author of this paper.

VITA

2000	Bachelor of Science, University of Delhi, India
2003	Master of Science, National Center for Biological Sciences, India
2004	Research Assistant, University of California, San Diego *Laboratory of Professor Steven A. Hillyard, Ph.D.*
2006	Teaching Assistant, Department of Biology University of California, San Diego
2008	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Mishra J, Hillyard SA. Endogenous attention selection during binocular rivalry at early stages of visual processing. Vision Research. *in press*. 2008.

Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze H, Hillyard SA. Neural basis of the Ventriloquist illusion. Current Biology: 17: 1-7. 2007.

Mishra J, Martinez A, Sejnowski TJ, Hillyard SA. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. Journal of Neuroscience 27: 4120-4131. 2007.

Mishra J, Fellous JM, and Sejnowski TJ. Selective attention through phase relationship of excitatory and inhibitory input synchrony in a model cortical neuron. Neural Networks 19: 1329-46. 2006.

Mishra J, and Bhalla US. Simulations of Inositol Phosphate Metabolism and its Interaction with $InsP_3$ mediated Calcium Release. Biophysical Journal 83: 1298-1316. 2002.

Sivakumaran S, Hariharaputran S, Mishra J, and Bhalla US. The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks. Bioinformatics 19: 408-415. 2003.

Kothekar V, Sahi S, Srinivasan M, Mohan A, and Mishra J. Recognition of cyclooxygenase-2 (COX-2) active site by NSAIDs: a computer modeling study. Indian Journal of Biochemistry & Biophysics 38: 56-63. 2001.

Kothekar V, Sahi S, and Mishra J. Molecular dynamics simulation of the interaction of 5-keto substituted 7-*tert*-butyl-2,3-dihydro-3,3-dimethylbenzofuran derivatives with cyclooxygenase-2. Current Science 80: 764-770. 2001.

Kothekar V, Sahi S, and Mishra J. Enzyme selectivity of new cyclooxygenase-2/5 lipoxygenase inhibitors using molecular modeling approach. Indian Journal of Biochemistry & Biophysics 37: 86-96. 2000.

ABSTRACT OF THE DISSERTATION

Neural Processes underlying an Auditory-induced Visual Illusion

by

Jyoti Mishra

Doctor of Philosophy in Biology / Specialization Computational Neurobiology

University of California, San Diego, 2008

Professor Steven Hillyard, Chair

Perception is multimodal in that it results from the integration of sensory inputs simultaneously provided by natural stimuli in multiple modalities. Investigations on how the brain combines information from different senses are key to understanding the mechanism of perception. Multisensory illusions are an interesting aspect of integration of the senses, wherein the percept in one modality is altered by the co-presentation of stimuli in another modality. The neural bases of such illusions can provide fundamental insights into the implementation of sensory integration in the brain, yet have not been investigated in much detail. One such illusion is the sound-induced extra flash illusion wherein a flash presented in conjunction with two pulsed sounds generates the percept of two flashes, of which the second is illusory (Shams et al., 2000, 2002). A thorough

analysis of the cortical mechanisms that underlie this striking audio-visual illusion is the focus of the present thesis.

The neural correlates of the sound-induced illusory extra flash were investigated using event related potential (ERP) recordings that provide high temporal resolution, along with anatomical source localization techniques. The sensory properties of the illusory flash were further explored in a behavioral study. The key findings revealed that the illusion results from a rapid dynamic interplay between processing in auditory and visual cortical areas in conjunction with activity within polymodal superior temporal cortex. This activation sequence included the finding of a novel ERP component, the PD120, that was elicited rapidly within 30-60 ms of the second sound in the illusion inducing audio-visual configuration. The amplitude of the PD120 was found to be strongly correlated with the frequency of illusory percepts in individual subjects. Source localization analyses in two separate studies, and the absence of polarity inversion of the PD120 component as a function of stimulus field location, demonstrated that the principal generator of this component lies outside of striate cortex, in the ventral occipito-temporal region of extrastriate visual cortex. It was also found that subjects who frequently perceived the illusion showed this early modulation (PD120) in response to other combinations of auditory and visual stimuli, thus pointing to consistent individual differences in the neural connectivity that underlies cross-modal integration. Attention was found to significantly enhance the cross-modal integration processes underlying the illusion and in particular was shown to be crucial to the generation of the PD120 component. Finally, a behavioral study found that the extra flash illusion is a robust

phenomenon that can be generated for visual stimulus features such as color and basic shape.

Overall the findings underscore the emerging view that multisensory integration involves a rapid dynamic interplay, rather than a serial progression, between unisensory cortices and the traditional cortical association areas such as the superior temporal region. These studies emphasize further that even though unisensory brain areas show preferential responsiveness to a specific modality, they are crucial to the process of multimodal perception.

REFERENCES

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

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147-152.

Chapter 1: Neural Processes underlying Multisensory Integration, and the generation of Cross-modal Illusions

INTRODUCTION

Humans possess multiple senses through which we perceive, experience and explore the natural environment. The external stimuli that impinge upon the senses most often have properties that provide information to more than one sensory modality. The study of multisensory integration provides insights into how information from the different senses is brought together in the brain to form unified and coherent percepts of the external world (Stein and Meredith, 1993, Calvert et al., 2004). This integration ability provides many behavioral advantages such as improved detection and discrimination capabilities as well as speeded reaction times in response to stimuli (Zahn et al., 1978, Stein et al., 1989, Perrott et al., 1990 Hughes et al., 1994, Frens et al., 1995, McDonald et al., 2000, Newell, 2004).

The study of cross-modal integration was pioneered by Stein and colleagues using the subcortical region of the superior colliculus (SC) as a model system (Stein and Arigbede, 1972, Stein, 1978, Meredith and Stein, 1983, 1986). Multisensory neurons were characterized in the deep layers of the SC that had strong responses to co-stimulation of two or more of the visual, auditory and somatosensory modalities. Fig. 1 demonstrates the non-linear super-additive response properties of one such multisensory neuron that had overlapping visual and auditory receptive fields. The firing rates of such multisensory neurons to combined visual and auditory stimulation greatly exceeded the sum of their firing rate responses to either stimulus presented alone.

In humans, multisensory associations have been studied using electrophysiological techniques such as event related potential (ERP) recordings and neuroimaging methods such as functional magnetic resonance imaging (fMRI) (Calvert, 2001). In some cases, transcranial magnetic stimulation (TMS) has been applied to impart transient lesions to brain sites to investigate their role in cross-modal association (Ruff et al., 2006, Romei et al., 2007). Within the cortex, predominant heteromodal regions reside in the superior temporal sulcus (STS), intraparietal sulcus (IPS) and within the ventrolateral prefrontal cortex (Driver and Noesselt, 2008). More recently, sensory cortical regions traditionally considered to be unimodal such as occipital visual cortex and temporal auditory cortex have also been implicated as sites of crossmodal integration (Foxe and Schroeder, 2005, Ghazanfar and Schroder, 2006).

Brain areas have been considered sites of cross-modal integration if they meet the nonlinear criterion that activation of these sites by bi- or tri-modal stimulation deviates from additive responses to the separate unimodal stimuli, similar to the response properties of SC neurons (Calvert and Thesen, 2004, Stein et al., 2004). The function of the different cross-modal regions depends on their recruitment within different cross-modal tasks/ cross-sensory phenomena. These include:

i) cross-modal matching, in which subjects compare and match features across the senses,
ii) cross-modal identification, which includes linguistic studies such as synthesis of audio-visual speech, as well as non-linguistic cross-modal object recognition,
iii) cross-modal localization, which is based on combining the spatial coordinates of information across the senses,

iv) cross-modal attention, which includes attention to multisensory stimuli, as well as studies of the spread of attention from the primary task modality to the secondary task modality,

v) cross-modal learning based on assimilation of newly paired information across the senses,

vi) cross-modal illusions, wherein the percept within one unisensory modality is significantly altered by coincident information from a different modality.

The focus of the current thesis lies within the last category, specifically, how visual perception is altered by audition. The following sections provide an overview of the key findings with respect to the functioning of different multisensory cortical convergence regions in the context of the above categories. Emerging models of the current understanding of multisensory interactions in the brain are discussed. Following the general review of the field, a specific background to cross-modal illusions and the sound-induced extra flash illusion, which forms the focus of this thesis, is provided.

Cross-modal information processing in different brain regions

Cross-modal Matching

Tasks wherein cross-modal matching is the primary goal have been predominantly explored in humans using functional neuroimaging such as PET (Positron Emission Tomography) and fMRI. In visuo-tactile matching studies, wherein subjects matched the shape of a tactile object to visual shapes on the screen, a network of multisensory brain areas was activated including the inferior parietal lobes, bilateral superior temporal sulcus (STS), the anterior cingulate and regions of prefrontal cortex,

along with robust activity within the insula-claustrum (Banati et al., 2000, Ettlinger and Wilson, 1990, Hadjikhani and Roland, 1998). Importantly, activity related to cross-modal matching was isolated as the residual activity after subtracting neural activity within control unisensory matching tasks. Paradigms that compared presentation of naturally congruous and incongruous audio-visual stimuli, but did not have matching as an explicit task demand, have also demonstrated activation within parts of the above network, such as the anterior cingulate (Laurienti et al., 2003) and the STS (Barraclough et al., 2005).

Cross-modal Object Recognition

In studies of cross-modal object recognition, the different senses provide complimentary information about the features of an object. It is being increasingly found that unimodal cortices interact in such tasks with or without information relay through typical multisensory sites (Amedi et al., 2005). For instance, in the case of audio-visual speech, super-additive effects were found in both auditory and visual cortices, apart from STS, when the speaker could be seen as well as heard (Calvert et al., 1999, 2000). More recently, fMRI of subjects watching simple lip reading (that is, visual speech alone) demonstrated above baseline auditory cortex activity in the Heschl's gyrus (Pekkola et al., 2005).

With respect to non-linguistic processing, an ERP study found that objects formed by arbitrary associations between auditory (tones) and visual (shape) features modulated both auditory and visual cortices along with heteromodal sites (Giard and Peronnet, 1999). However, in this and some subsequent ERP studies with similar findings (Foxe et al., 2000, Molholm et al., 2002), cross-modal interactions within unisensory cortices were

suggested to occur at latencies of less than 100 ms post-stimulus onset. Teder-Salejarvi et al. (2002) showed that these very early cross-modal interactions were most probably an artifact of contamination of the early post-stimulus period by overlapping anticipatory pre-stimulus activity. Many multisensory studies in recent years have taken precaution against such errors by taking into account neuronal activity elicited by a blank stimulus or rest period in the calculation of cross-modal interactions (Calvert et al., 2001, Teder-Salejarvi et al., 2002, Talsma and Woldorff, 2005, Gondan and Roder, 2006). Interaction effects occurring later than 100 ms after stimulus onset (in the range of the visual N1 ERP component) that were related to processing of matching object pictures and sounds, were shown to localize to lateral occipital cortex (LOC) in the ventral visual stream (Molholm et al., 2004). Overall, the superior temporal sulcus and gyrus (STS/ STG) region has been consistently found to play a critical role in cross-modal object identification. Moreover, activity within this region is strongest when sensory information is degraded by noise (Callan et al., 2001, 2003, Sekiyama et al., 2003) in accordance with the *principle of inverse effectiveness*, which is commonly exhibited by superior collicular neurons (Stein and Meredith, 1993). This principle refers to the phenomenon that multisensory responses in SC neurons are largest when each unisensory input alone elicits a relatively weak response, as for less intense stimuli.

Spatial Congruity in Cross-modal Processing

One of the first imaging studies delineating cortical sites of cross-modal co-localization was a visuo-tactile fMRI study (Macaluso et al., 2000). The most robust activation to spatial concordance across modalities was found in the inferior parietal lobe

(IPL), but modulations were also seen in the lingual gyrus of the visual cortex. Although fMRI cannot provide any temporal information, it was hypothesized that activity within occipital areas was a result of feedback from IPL, which is known to process visuo-spatial information (Macaluso and Driver, 2001, Spence and Driver, 2004). An ERP study of audio-visual spatial congruity also found differential modulation within ventral occipital cortex to congruent vs. incongruent stimuli 100-400 ms post stimulus onset, along with an amplitude modulation within STS at around 250 ms (Teder-Salejarvi et al., 2005). The involvement of different cross-modal regions, IPL vs. STS in visuo-tactile vs. audio-visual paradigms, respectively, may be in line with a recent proposal that cross-modal integration occurs in multisensory zones at the borders between the involved unisensory regions (Wallace et al., 2004).

Attention and Cross-modal Integration

Cross-modal attention has been investigated in the context of space, attending to a cross-modal stimulus in one spatial location while ignoring stimuli at a different location, as well as with respect to modality, that is attending to information presented in one of two modalities while ignoring the other. Using ERPs, Hillyard and colleagues (Hillyard et al., 1984, Teder-Salejarvi et al. 1999), showed that processing of stimuli at attended locations in space is facilitated at an early, sensory level, for both the attended as well as the unattended modality. In these studies subjects attended to either auditory or visual stimuli in the left or right hemifield. When auditory stimuli were attended, the auditory N1 evoked to the auditory stimulus was enhanced within the attended hemifield relative to unattended trials when attention was directed in the opposite hemifield. A cross-modal

effect was also observed as the visual N1 evoked by unattended visual stimuli at the location of the attended auditory stimulus showed a location-specific attention effect. The complementary effect was demonstrated for unattended auditory stimuli when co-localized visual stimuli were attended within a hemifield. Of note, in all cases the intra-modal attention effects were larger than the cross-modal attention effects. These and other parallel studies (Spence and Driver, 1996, Eimer and Schroger, 1998, Eimer et al., 2001, 2002, Talsma and Kok, 2002, Ciaramitaro et al., 2007) highlight the multi-modal organization of spatial attention, such that attention allocated to a given spatial location in a primary modality spreads across to other modalities in an attenuated manner (Driver and Spence, 1998).

With respect to involuntary attention, McDonald et al. (2003) showed that spatial attention driven by sounds enhanced visual evoked potentials to flashes at the sound's location. Neural activity was found to be modulated in the region of the STG at 120-140 ms rapidly followed by a modulation within visual extrastriate fusiform gyrus. Owing to the high temporal resolution of ERPs, the effect in visual cortex was inferred to be a result of feedback from multisensory area STG. A corresponding effect on auditory neural responses was also found for attention to audition driven by visual cues (McDonald et al., 2001).

Attention to bimodal audio-visual stimuli has been shown to significantly affect their neural processing at multiple stages. In an ERP study, all components evoked by multisensory stimuli from 100-500 ms were significantly enhanced by attention (Talsma and Woldorff, 2005). Talsma et al. (2007) also demonstrated very early attention affects, at 50 ms post-stimulus, when both visual and auditory senses were simultaneously

attended but not when attention was focused to just one of the two modalities. Studies of attention effects on bimodal stimulus processing, although few, seem to suggest that the early sensory gain control mechanism of attention that applies to unisensory processing (Hillyard et al., 1998) may also extend to multisensory inputs.

Cross-modal Learning

Perceptual learning in the visual modality has been shown to be improved when learning is induced through audio-visual training (Seitz et al., 2006, Kim et al., 2008). For example, such a facilitation in learning was shown for auditory voice recognition following paired voice-face associations (vonKriegstein and Giraud, 2006). In neuroimaging studies, it has been found that development of cross-modal associations by constantly pairing bimodal stimuli, recruits neural activity in unisensory cortices. Specifically, if one stimulus of the cross-modal pair is presented in isolation following the pairing, the cross-modal unisensory cortex is activated (McIntosh et al., 1998, Calvert, 2001, Baier et al., 2006). Similar effects have also been found for cross-modal associations learned via experience; for example, lip reading of visual speech was shown to activate auditory cortex (Pekkola et al., 2005). Also worth mentioning, Murray et al. (2004) showed in an ERP study that neural responses evoked by visual stimuli previously learned in association with sounds versus without sounds, generated a modulation of lateral occipital cortex as early as 60-140 ms.

Models of Cross-modal Integration

The above studies clearly demonstrate that the brain areas involved in multisensory integration cannot be assigned exclusive roles but are recruited as part of a network in varying ways in order to extract useful stimulus information for successful task performance. A model demonstrating the interplay between brain areas, while suggesting broad roles for various participant regions is shown in Fig. 2 (adapted from Calvert, 2001). Again, to emphasize, even though the figure suggests functional specializations of different areas to aid understanding, such specializations are far from the rule. At the other extreme of this argument, it has been suggested that the entire neo-cortex may be multisensory (Ghazanfar and Schroeder, 2006). The above studies, however, do not suggest complete non-differentiation, but emphasize varying degrees of involvement between the traditional multisensory and unisensory cortices in different cross-modal phenomena (Driver and Noesselt, 2008).

Another model supported by recent findings (Beauchamp et al., 2004, Wallace et al., 2004, Cappe and Barone, 2005) posits that multisensory convergence zones may extend into sensory-specific cortices further than classical evidence has suggested. For instance, in a detailed mapping study in rats (Wallace et al., 2004), multisensory neurons were consistently found near the borders of unisensory cortices. The response profiles of these neurons integrated inputs from the sensory modalities that lay on either side of the border. In macaques the belt region of auditory cortex that surrounds the unisensory core region has been shown to receive somatosensory input in feedforward layers (layer 4), while visual input projects to the same region within feedback layers 2 and 3 (Schroeder and Foxe, 2002, 2005, Kayser et al., 2005). In a recent study, sparse visual projections

into primary auditory cortex A1 were also found, but the majority of audio-visual convergence zones were traced to higher auditory areas (Bizley et al., 2007). In visual cortex, auditory input has been shown to project to areas as early as V1 and V2 using anatomical labeling (Falchier et al., 2002, Rockland and Ojima, 2003, Clavagnier et al., 2004). These connections are sparse, however, and project predominantly to regions that encode peripheral rather than central visual field inputs.

Contrasting with the above models, another view of cross-modal integration emphasizes feedback from polymodal areas to unisensory cortices as being critical to multisensory information processing (Driver and Noesselt, 2008). The results of some studies described in the above section, such as the visuo-tactile spatial co-localization study by Macaluso et al. (2000) and the sound induced cross-modal attention effects on visual processing described by McDonald et al. (2003), do fit this framework. An fMRI study investigating the correlates of temporal coincidence between audiovisual stimuli also inferred feedback from STS to primary areas A1 and V1 via a connectivity analysis (Noesselt et al., 2007). However, it is evident from the majority of the studies reviewed here that feedback is not a universal mechanism of cross-modal integration.

Overall, the above described models implicate a rich interplay between brain regions contrary to the traditionally proposed scheme wherein cross-modal interaction occurred exclusively in polymodal association cortex subsequent to processing of unimodal components of the input in unisensory cortices. This accords with an emerging view that sensory processing involves parallel and recursive processing loops (Kaas and Hackett, 2005, Scheich et al., 2007) rather than simple serial progression (Driver and Noesselt, 2008).

Cross-modal Illusions

Multisensory illusions are an intriguing subset of cross-modal phenomena. Behaviorally, cross-modal integration has been demonstrated to enhance perception by improving detection and discrimination abilities and to speed reaction times. In the case of illusions, however, the percept in one sensory modality is altered by the co-occurrence of events in another modality. In some cases, in fact, the percept of the cross-modal stimulus maybe completely different from its unimodal parts. Popular examples of such phenomena are the ventriloquist illusion and the McGurk effect. The former is a ubiquitous phenomenon in everyday experience, wherein the spatial location of a sound source is illusively pulled to the location of the visual stimulus associated with the sound (Bertelson and Radeau, 1981, Bertelson, 1999). In the McGurk effect, when auditory and visual speech stimuli corresponding to different syllables are co-presented (e.g. audio /ga/ and visual /ba/), the audiovisual percept is of an altogether new speech sound (/da/) (McGurk and MacDonald, 1976). In a sense, such illusions are cases of *mis*-integration or conflict between the senses, but their existence suggests important principles that the brain may follow to generate optimal perception, which can be uncovered by studying such phenomena.

The neurophysiological basis of some of these illusions is only beginning to be understood. Bonath et al. (2007) recently combined ERP recordings and fMRI to show that the illusory spatial displacement of sounds by visual stimuli in the ventriloquist effect is tracked on a trial-by trial basis by the relative modulation of left and right auditory cortices. On trials when tones presented from a central speaker were perceived

to be located in the left (or right) field where a flash occurred (about 33% of the trials on average), a precisely timed component in the multimodal interaction waveforms measured at 230-270 ms (N260) showed a voltage distribution that was shifted over to the hemisphere contralateral to the visual stimulus. (The multimodal interaction waveforms were calculated by subtracting the sum of ERPs on separate auditory and visual trials from the ERPs on audio-visual trials.) On non-illusory trials when subjects correctly perceived the location of the sound source as central, the scalp distribution of the N260 component was balanced between the hemispheres. Dipole source localization confirmed the N260 to arise from neural generators within auditory cortex. Converging results were obtained with fMRI, with auditory cortex activity on illusion trials being lateralized to the hemisphere contralateral to the ventriloquized location of the sound. Thus, although previous physiological studies have used audiovisual paradigms with spatially disparate auditory and visual stimuli that could induce ventriloquism (Busse et al., 2005, Gondan et al., 2005, Teder-Salejarvi et al., 2005, Bischoff et al., 2007), this was the first study to show that the neural correlate of ventriloquism was a lateralized neural interaction in auditory cortex.

Mottonen et al. (2002) used magnetoencephalographic (MEG) recordings to study the neural basis of the McGurk effect using Finnish audiovisual speech stimuli. Auditory /ipi/ sounds were co-presented with congruent visual /ipi/ standards or incongruent deviant /iti/ articulations. Audio /ipi/ with visual /iti/ were together perceived as /iti/. Congruent /iti/ deviants (both audio and visual /iti/) were also presented on a small percentage of trials. Visual only trials with either /ipi/ standards or deviant /iti/ articulations were also presented for comparison to audiovisual trials. The authors found

that deviant visual speech stimuli in audiovisual presentations produced a bilateral modulation within auditory cortex as early as 100 ms. The deviant visual /iti/ stimuli in unimodal presentations also produced a modulation within auditory cortex, but at a longer latency than the same deviants within the audiovisual presentations. The authors suggested that the cross-modal integration underlying the McGurk effect occurred at an early stage in auditory cortex, well before the initiation of phonetic classification.

Cross-modal Influences on Vision

For both of the common cross-modal illusions described above, vision predominates over audition. Although vision is largely considered the dominant modality in humans, even visual perception can be altered by other senses such as audition. Most such perceptual changes have been reported with respect to temporal processing (Gebhard and Mowbray, 1959, Shipley, 1964, Fendrich and Corballis, 2001, Aschersleben and Bertelson, 2003, Berger et al., 2003, Morein-Zamir et al., 2003, Recanzone, 2003, Vroomen and deGelder, 2004). In fact, the term '*temporal ventriloquism*' has been coined for such phenomena as when the perceived rate of a flickering visual stimulus is strongly influenced by a concurrent repetitive auditory stimulus presented at a very different temporal rate. Temporal ventriloquism has even been shown to share similarities with conventional ventriloquism in that they both leave aftereffects on the influenced modality. Thus, in ventriloquism the spatial location of the auditory stimulus remains shifted in the direction of a visual stimulus for a while even after the visual stimulus has been removed (Lewald, 2002), and in temporal ventriloquism the flicker rate of a repetitive visual stimulus remains altered for a brief

period after removal of the temporally disparate auditory stimulus (Recanzone et al., 2003).

Audition can shift the temporal order judgment (TOJ) of co-occurring visual stimuli. When visual stimuli (light flashes) are co-presented at two different locations following a sound at one of the two locations, the flash coincident with the location of the sound is perceived to occur significantly earlier than the flash at the different location. McDonald et al (2005) showed that the temporal lead of the flash influenced by the prior sound at the same location can be as large as 70 ms. An intriguing question was whether the physiological processing time of the earlier visual stimulus would also be hastened akin to the behavioral effect. Using ERPs McDonald et al. (2005) showed that the auditory cueing did not affect the latency of the contralateral cortical response to the first perceived flash relative to the second flash. The cueing, rather, enhanced the amplitude of neural activity in visual cortex to the first flash relative to the second. The enhanced signal was observed in the ventral processing stream of extrastriate visual cortex onsetting at 80 ms and overlapping the principal visual evoked P1, N1 and P2 components. This finding demonstrated that cross-modally induced shifts in visual time-order perception can arise from modulations of signal strength rather than processing speed in the early visual cortical pathways. These results also underscore the importance of investigating the neural bases of such phenomena, as the neural mechanisms can turn out to be different from those intuitively suggested by the behavior.

Hypothesis of 'Modality Appropriateness'

Overall, the above phenomena suggest the dominance of vision in the spatial domain and that of audition in the temporal domain. The 'modality appropriateness hypothesis' is a theoretical proposal that attempts to explain these observations by postulating that the modality which provides the more precise and less variable information about the task at hand dominates perception (Welch et al., 1986). Cross-modal influences are thus suggested to be a result of optimal integration between sensory inputs that takes into account the variability of the incoming information (Witten and Knudsen, 2005). Bayesian models have been successfully employed to demonstrate how such integration is statistically feasible. The cross-modal output in these models is generated as a weighted sum of the inputs from the involved modalities, where the weight for each modality is inversely proportional to the variance (or directly proportional to the reliability) of that input (Ernst and Banks, 2002, Kording et al., 2007). How the brain actually implements such a theoretical principle can be elucidated by studying the neural correlates of cross-modal illusions.

The focus of the present thesis is a qualitative change in visual perception induced by audition. A couple of examples of such phenomena have been described in the literature. Stein et al. (1996) showed that the perceived intensity of a flash at fixation is enhanced by the concurrent presence of a sound. Sekuler et al. (1997) showed that two visual object moving towards each other in a collision path normally appear to pass each other and continue motion in their original paths, but in the presence of a sound coinciding with the visual stimuli, the perception of the interaction between the two visual objects dramatically changed to a collision and rebounding of the objects.

The Sound-Induced Extra Flash Illusion

A striking change in visual perception induced by audition was recently described by Shams et al. (2000, 2002) and termed "the sound induced illusory flash" phenomenon. In this phenomenon, presentation of a single brief flash interposed between two pulsed sounds separated by 60-100 ms typically results in the perception of two distinct flashes, of which the second is illusory.

The neural basis of this illusory extra flash percept was initially investigated through ERP recordings. Shams et al. (2001) isolated neural activity corresponding to the perceived second flash by calculating an ERP difference wave in which ERPs to the constituent unimodal auditory and visual stimuli were subtracted from the ERP to the illusion-inducing auditory-visual-auditory ensemble. A positive deflection in this difference ERP over the occipital scalp peaking at around 200 ms was taken as evidence for visual cortex involvement in perceiving the illusory flash. Arden et al. (2003) made a similar analysis but found a much earlier negative deflection (at 100-125 ms after the first sound) in the occipital difference ERP, interpreted as a rapid modulation of visual cortex by the second sound. Since ERPs were recorded from only a few electrodes in these studies, however, it was not possible to verify the neural origins of these difference wave components. In a subsequent study using magnetoencephalography (MEG), Shams et al. (2005a) reported an even earlier difference wave modulation (latency 35-65 ms) at right occipital sensors, which began even prior to presentation of the second sound. Given these inconsistencies in the ERP/MEG results to date, it is not clear which neural events may be critical for producing this sound-induced illusory percept.

In a recent fMRI investigation of the phenomenon, Watkins et al. (2006) reported selective enhancement of cortical activity in primary visual cortex (area V1) and in certain polysensory areas on trials on which the illusion was perceived. Given the low temporal resolution of fMRI, however, the timing of these effects with respect to the auditory-visual stimulation could not be determined. It was also unclear whether the cortical activation pattern associated with the illusory flash percept was comparable to that associated with perception of an actual flash.

EXPERIMENTAL AIMS

The present thesis aims to characterize the cortical dynamics that underlie the illusory flash phenomenon and to gain insight into the neural mechanisms that govern this altered perceptual experience. The neural basis of the cross-modal illusion was investigated using 64-channel ERP recordings in conjunction with anatomical source localization. The specific studies carried out are outlined below:

i) Using high temporal resolution ERP recordings the spatio-temporal neural activity pattern specifically associated with perception of the illusory second flash was detailed in Chapter 2. This activity profile was also compared with the activity elicited by a real second flash.

ii) Complementary to the two-sound-one-flash stimulus that results in the visual perception of two flashes, it has been observed that one sound in conjunction with two flashes generates the percept of a single flash (Andersen et al., 2004, Shams et al., 2005b). Using ERPs along with source localization analyses, the neural basis of this flash fusion phenomenon was investigated in Chapter 3.

iii) The focus of Chapter 4 was to explore the role of selective attention in generation of the extra flash illusion. Audiovisual stimuli were presented in both the upper and lower visual fields in random order. Subjects selectively attended to one of the two stimulus locations and reported visual percepts at that location, while ignoring the other location. The influence of attention on the cross-modal ERPs underlying the illusory flash percept was studied.

iv) In Chapter 5 the sensory properties of the sound-induced illusory extra flash percept was investigated in a behavioral study. It was explored whether the illusion can affect visual stimuli with specific features such as color and shape. It was also studied how the illusory flash interacts with a sequence of two real flashes and whether it can affect the temporal association between these real flashes.

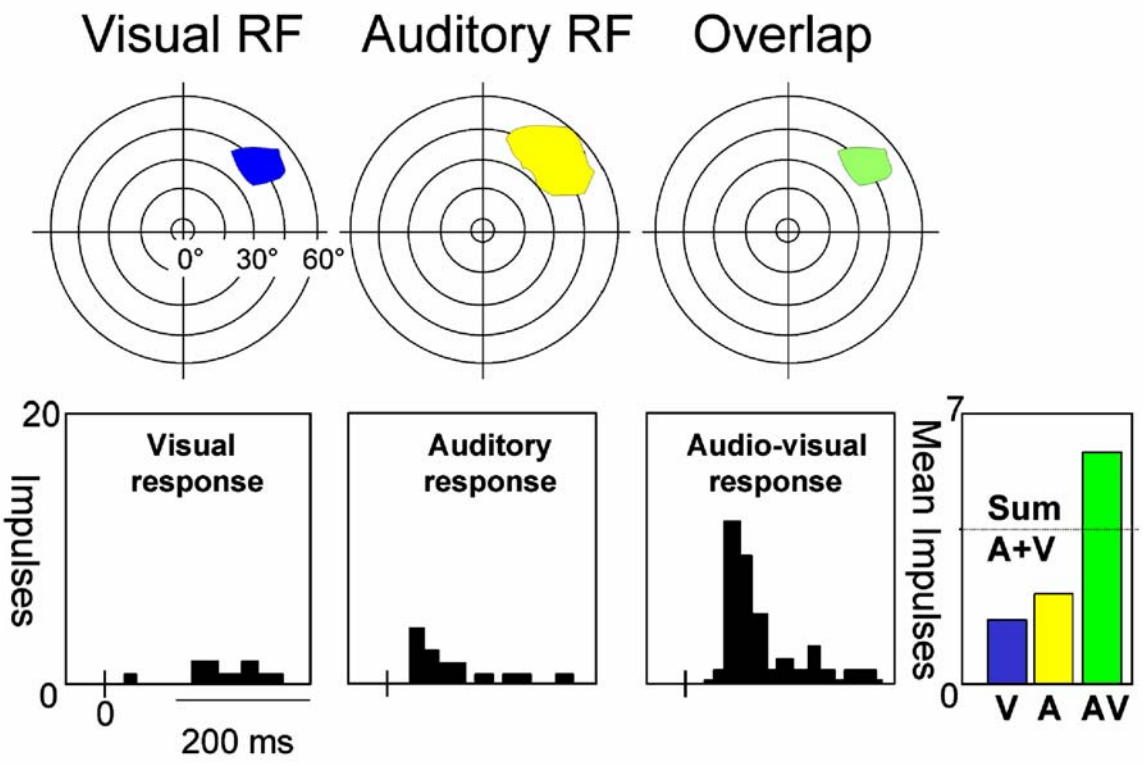

Figure 1.1 Response properties of a multisensory neuron in deep superior colliculus. The neuron has overlapping visual and auditory spatial receptive fields (RF). The bottom row illustrates that the neuron's firing response to audio-visual (AV) stimuli is super-additive to the sum of its responses to each modality alone (adapted from Driver and Noesselt, 2008).

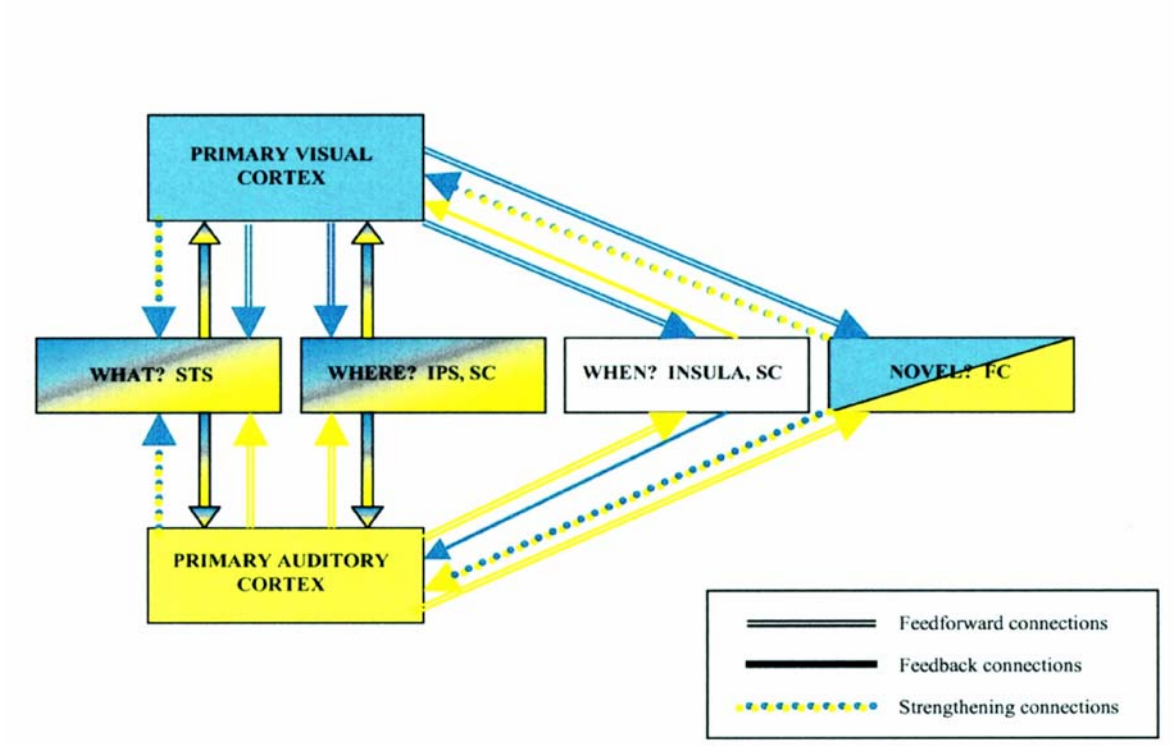

Figure 1.2 Cross-modal integration results from an interplay between different brain regions. The illustration is in context to audiovisual integration. STS, IPS, SC and FC refer to the Superior Temporal Sulcus, Inferior Parietal Sulcus, Superior Colliculus and Frontal Cortex, respectively (adapted from Calvert, 2001).

REFERENCES

Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. Exp Brain Res 166:559-571.

Andersen TS, Tiippana K, Sams M (2004) Factors influencing audiovisual fission and fusion illusions. Brain Res Cogn Brain Res 21:301-308.

Arden GB, Wolf JE, Messiter C (2003) Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. Vision Res 43:2469-2478.

Aschersleben G, Bertelson P (2003) Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. Int J Psychophysiol 50:157-163.

Baier B, Kleinschmidt A, Muller NG (2006) Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. J Neurosci 26:12260-12265.

Banati RB, Goerres GW, Tjoa C, Aggleton JP, Grasby P (2000) The functional anatomy of visual-tactile integration in man: a study using positron emission tomography. Neuropsychologia 38:115-124.

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17:377-391.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 7:1190-1192.

Berger TD, Martelli M, Pelli DG (2003) Flicker flutter: is an illusory event as good as the real thing? J Vis 3:406-412.

Bertelson P (1999) Ventriloquism: A case of crossmodal perceptual grouping. In: Cognitive contributions to the perception of spatial and temporal events. (Achersleben G, Bachman T, Musseler J, eds). Amsterdam: Elsevier.

Bertelson P, Radeau M (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Percept Psychophys 29:578-584.

Bischoff M, Walter B, Blecker CR, Morgen K, Vaitl D, Sammer G (2007) Utilizing the ventriloquism-effect to investigate audio-visual binding. Neuropsychologia 45:578-586.

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17:2172-2189.

Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA (2007) Neural basis of the ventriloquist illusion. Curr Biol 17:1697-1703.

Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci U S A 102:18751-18756.

Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001) Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Brain Res Cogn Brain Res 10:349-353.

Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E (2003) Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport 14:2213-2218.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110-1123.

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport 10:2619-2623.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649-657.

Calvert GA, Stein BE, Spence C (2004) The handbook of multisensory processing. Cambridge, MA: MIT.

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Paris 98:191-205.

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886-2902.

Ciaramitaro VM, Buracas GT, Boynton GM (2007) Spatial and cross-modal attention alter responses to unattended sensory information in early visual and auditory human cortex. J Neurophysiol 98:2399-2413.

Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. Cogn Affect Behav Neurosci 4:117-126.

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57:11-23.

Driver J, Spence C (1998) Crossmodal attention. Curr Opin Neurobiol 8:245-253.

Eimer M, Cockburn D, Smedley B, Driver J (2001) Cross-modal links in endogenous spatial attention are mediated by common external locations: evidence from event-related brain potentials. Exp Brain Res 139:398-411.

Eimer M, Schroger E (1998) ERP effects of intermodal attention and cross-modal links in spatial attention. Psychophysiology 35:313-327.

Eimer M, van Velzen J, Driver J (2002) Cross-modal interactions between audition, touch, and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations. J Cogn Neurosci 14:254-271.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429-433.

Ettlinger G, Wilson WA (1990) Cross-modal performance: behavioral processes, phylogenetic considerations and neural mechanisms. Behav Brain Res 40:169-192.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749-5759.

Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Percept Psychophys 63:719-725.

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res 10:77-83.

Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16:419-423.

Frens MA, Van Opstal AJ, Van der Willigen RF (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Percept Psychophys 57:802-816.

Gebhard JW, Mowbray GH (1959) On discriminating the rate of visual flicker and auditory flutter. Am J Psychol 72:521-529.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278-285.

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473-490.

Gondan M, Niederhaus B, Rosler F, Roder B (2005) Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. Percept Psychophys 67:713-726.

Gondan M, Roder B (2006) A new method for detecting interactions between the senses in event-related potentials. Brain Res 1073-1074:389-397.

Hadjikhani N, Roland PE (1998) Cross-modal transfer of information between the tactile and the visual representations in the human brain: A positron emission tomographic study. J Neurosci 18:1072-1084.

Hillyard SA, Anllo-Vento L (1998) Event-related brain potentials in the study of visual selective attention. Proc Natl Acad Sci U S A 95:781-787.

Hillyard SA, Simpson GV, Woods DL, VanVoorhis S, Munte TF (1984) Event-related brain potentials and selective attention to different modalities. In: Cortical Integration (Reinoso-Suarez F, Aimone-Marsan C, eds), pp 395-413. New York: Raven Press.

Hughes HC, Reuter-Lorenz PA, Nozawa G, Fendrich R (1994) Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. J Exp Psychol Hum Percept Perform 20:131-153.

Kaas J, Hackett T (2005) Subdivisions and connections of the auditory cortex in primates. A working model. In: The Auditory Cortex - A Synthesis of Human and Animal Research (Konig R, Heil E, Budinger E, Scheich H, eds), pp 7-26. Mahwah, NJ: Lawrence Erlbaum.

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48:373-384.

Kim RS, Seitz AR, Shams L (2008) Benefits of stimulus congruency for multisensory facilitation of visual learning. PLoS ONE 3:e1532.

Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS ONE 2:e943.

Laurienti PJ, Wallace MT, Maldjian JA, Susi CM, Stein BE, Burdette JH (2003) Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. Hum Brain Mapp 19:213-223.

Lewald J, Guski R (2003) Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. Brain Res Cogn Brain Res 16:468-478.

Macaluso E, Driver J (2001) Spatial attention and crossmodal interactions between vision and touch. Neuropsychologia 39:1304-1316.

Macaluso E, Frith CD, Driver J (2000) Modulation of human visual cortex by crossmodal spatial attention. Science 289:1206-1208.

McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2003) Neural substrates of perceptual enhancement by cross-modal spatial attention. J Cogn Neurosci 15:10-19.

McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2005) Neural basis of auditory-induced shifts in visual time-order perception. Nat Neurosci 8:1197-1202.

McDonald JJ, Teder-Salejarvi WA, Heraldez D, Hillyard SA (2001) Electrophysiological evidence for the "missing link" in crossmodal attention. Can J Exp Psychol 55:141-149.

McDonald JJ, Teder-Salejarvi WA, Hillyard SA (2000) Involuntary orienting to sound improves visual perception. Nature 407:906-908.

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746-748.

McIntosh AR, Cabeza RE, Lobaugh NJ (1998) Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. J Neurophysiol 80:2790-2796.

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389-391.

Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol 56:640-662.

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452-465.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115-128.

Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. Brain Res Cogn Brain Res 17:154-163.

Mottonen R, Krause CM, Tiippana K, Sams M (2002) Processing of changes in visual speech in the human auditory cortex. Brain Res Cogn Brain Res 13:417-425.

Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schnider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage 21:125-135.

Newell FN (2004) Cross-modal object recognition. In: The handbook of multisensory processes. (Calvert GA, Spence C, Stein BE, eds), pp 123-139. Cambridge, MA: MIT Press.

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J Neurosci 27:11431-11441.

Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport 16:125-128.

Perrott DR, Saberi K, Brown K, Strybel TZ (1990) Auditory psychomotor coordination and visual search performance. Percept Psychophys 48:214-226.

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078-1093.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19-26.

Romei V, Murray MM, Merabet LB, Thut G (2007) Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. J Neurosci 27:11465-11472.

Ruff CC, Blankenburg F, Bjoertomt O, Bestmann S, Freeman E, Haynes JD, Rees G, Josephs O, Deichmann R, Driver J (2006) Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. Curr Biol 16:1479-1488.

Scheich H, Brechmann A, Brosch M, Budinger E, Ohl FW (2007) The cognitive auditory cortex: task-specificity of stimulus representations. Hear Res 229:213-224.

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. Curr Opin Neurobiol 15:454-458.

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res Cogn Brain Res 14:187-198.

Seitz AR, Kim R, Shams L (2006) Sound facilitates visual learning. Curr Biol 16:1422-1427.

Sekiyama K, Kanno I, Miura S, Sugita Y (2003) Auditory-visual speech perception examined by fMRI and PET. Neurosci Res 47:277-287.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308.

Shams L, Iwaki S, Chawla A, Bhattacharya J (2005a) Early modulation of visual cortex by sound: an MEG study. Neurosci Lett 378:76-81.

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

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147-152.

Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. Neuroreport 12:3849-3852.

Shams L, Ma WJ, Beierholm U (2005b) Sound-induced flash illusion as an optimal percept. Neuroreport 16:1923-1927.

Shipley T (1964) Auditory Flutter-Driving of Visual Flicker. Science 145:1328-1330.

Spence C, Driver J (1996) Audiovisual links in endogenous covert spatial attention. J Exp Psychol Hum Percept Perform 22:1005-1030.

Spence C, Driver J (2004) Crossmodal Space and Crossmodal Attention. Oxford: Oxford University Press.

Stein BE (1978) Development and organization of multimodal representation in cat superior colliculus. Fed Proc 37:2240-2245.

Stein BE, Arigbede MO (1972) Unimodal and multimodal response properties of neurons in the cat's superior colliculus. Exp Neurol 36:179-196.

Stein BE, London R, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci 8:497-506.

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge, MA: MIT.

Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. J Cogn Neurosci 1:12-24.

Stein BE, Stanford TR, Wallace MT, Vaughan JW, Jiang W (2004) Crossmodal spatial interactions in subcortical and cortical circuits. In: Crossmodal Space and Crossmodal Attention (Spence C, Driver J, eds), pp 25-50: Oxford University Press.

Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex 17:679-690.

Talsma D, Kok A (2002) Intermodal spatial attention differs between vision and audition: an event-related potential analysis. Psychophysiology 39:689-706.

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098-1114.

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098-1114.

Teder-Salejarvi WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. J Cogn Neurosci 17:1396-1409.

Teder-Salejarvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106-114.

Teder-Salejarvi WA, Munte TF, Sperlich F, Hillyard SA (1999) Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. Brain Res Cogn Brain Res 8:327-343.

von Kriegstein K, Giraud AL (2006) Implicit multisensory associations influence voice recognition. PLoS Biol 4:e326.

Vroomen J, de Gelder B (2004) Temporal ventriloquism: sound modulates the flash-lag effect. J Exp Psychol Hum Percept Perform 30:513-518.

Wallace MT, Ramachandran R, Stein BE (2004) A revised view of sensory cortical parcellation. Proc Natl Acad Sci U S A 101:2167-2172.

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. Neuroimage 31:1247-56.

Welch RB, DuttonHurt LD, Warren DH (1986) Contributions of audition and vision to temporal rate perception. Percept Psychophys 39:294-300.

Witten IB, Knudsen EI (2005) Why seeing is believing: merging auditory and visual worlds. Neuron 48:489-496.

Zahn JR, Abel LA, Dell'Osso LF (1978) Audio-ocular response characteristics. Sens Processes 2:32-37.

Chapter 2: Early Cross-modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion

ABSTRACT

When a single flash of light is presented interposed between two brief auditory stimuli separated by 60-100 ms, subjects typically report perceiving two flashes (Shams et al., 2000, 2002). We investigated the timing and localization of the cortical processes that underlie this illusory flash effect in 34 subjects by means of 64 channel recordings of event-related potentials (ERPs). A difference ERP calculated to isolate neural activity associated with the illusory second flash revealed an early modulation of visual cortex activity at 30-60 ms after the second sound, which was larger in amplitude in subjects who saw the illusory flash more frequently. These subjects also showed this early modulation in response to other combinations of auditory and visual stimuli, thus pointing to consistent individual differences in the neural connectivity that underlies cross-modal integration. The overall pattern of cortical activity associated with the cross-modally induced illusory flash, however, differed markedly from that evoked by a real second flash. A trial-by-trial analysis showed that short-latency ERP activity localized to auditory cortex and polymodal cortex of the temporal lobe, concurrent with gamma bursts in visual cortex, were associated with perception of the double flash illusion. These results provide evidence that perception of the illusory second flash is based on a very rapid dynamic interplay between auditory and visual cortical areas that is triggered by the second sound.

INTRODUCTION

Our sensory systems are interconnected so as to integrate stimuli in different modalities and thereby achieve unified and coherent percepts of environmental events. Recent investigations of multisensory integration suggest cross-modal interactions occur not only in polysensory brain regions but in unisensory cortical areas as well (Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006; Macaluso, 2006; Martuzzi et al., 2006). Human event related potential (ERP) recordings have demonstrated that unisensory areas can be engaged in cross-modal processing both at very early as well as late time periods after stimulus onset (Giard and Peronnet, 1999; Molholm et al., 2002; Talsma et al., 2007, Murray et al., 2005; Meylan and Murray, 2007). Furthermore, the cross-modal interactions in these brain regions can be modulated by various factors such as temporal and spatial congruence of stimuli, extent of content association and attention (Calvert and Thesen, 2004; Busse et al, 2005; Teder-Sälejärvi et al., 2005; Baier et al., 2006; Johnson and Zatorre, 2006). Thus the emerging brain model of multisensory integration is of a dynamic and highly interactive network of brain regions.

For some types of cross-modal integration, the perception of a stimulus in one modality is altered by the occurrence of a stimulus in another modality. Numerous studies have shown, for example, that perception of a visual event can be modified by the presence of a concurrent sound (Stein et al., 1996; Sekuler et al., 1997; McDonald et al., 2003, 2005; Vroomen and Gelder, 2004; Fendrich and Corballis, 2001; Recanzone, 2003). A particularly striking perceptual alteration was recently described by Shams and colleagues (2000, 2002), who found that presenting a single brief flash interposed between two pulsed sounds separated by 60-100 ms typically results in the perception of

two distinct flashes. Investigating the neural basis of this double-flash illusion provides a powerful approach for revealing how information from different modalities is integrated in the brain. Moreover, since the illusion consists of a discrete visual perceptual event that varies on a trial-by-trial basis, it offers the possibility of isolating the critical sequence of neural events by which an auditory input induces a visual percept.

Prior ERP/MEG (Shams et al., 2001, 2005a; Arden et al., 2003) and fMRI investigations (Watkins et al., 2006) of the neural basis of the double-flash illusion have suggested that visual cortex activation underlies the perception of the illusory second flash. However, the exact timing of this visual cortex activity and the participation of other brain regions in engendering the illusion still remain unclear. The present study investigated the neural basis of the cross-modal double-flash illusion using 64-channel ERP recordings in conjunction with anatomical source localization. The aim was to define the sequence of dynamic cross-modal interactions underlying the sound-induced illusory percept, and thus reveal the interplay between different cortical areas that leads to the altered perceptual experience. The spatio-temporal activity pattern associated with perception of the illusory second flash was also compared with activity elicited by a real second flash to evaluate whether these processes shared any similarities.

MATERIALS & METHODS

Task and Stimuli

Thirty-four right-handed healthy adults (18 females, mean age 23.9 yrs) participated in the study after giving written informed consent as approved by the University of California, San Diego Human Research Protections Program. Each participant had normal or corrected-to-normal vision and normal hearing.

The experiment was conducted in a sound-attenuated chamber having a background sound level of 32 dB and a background luminance of 2 cd/m^2. Subjects maintained fixation on a central cross positioned at a viewing distance of 120 cm. Auditory (A) and visual (V) stimuli were delivered from a speaker and red light emitting diode (LED), respectively, both positioned 20º of visual angle to the left of fixation (Fig. 1A). The stimuli were presented laterally because the double flash illusion is reportedly accentuated in the visual periphery (Shams et al., 2002). Each visual stimulus was a 5 ms 75 cd/m^2 flash and each auditory stimulus was a 10 ms 76 dB noise burst. Nine different stimulus combinations were presented in random order on each block of trials (Fig. 1B). These included unimodal auditory stimuli, occurring singly (A_1) or in pairs (A_1A_2) and unimodal visual stimuli occurring singly (V_1) or in pairs (V_1V_2). Bimodal stimulus combinations included A_1V_1, $A_1V_1A_2V_2$, $A_1V_1A_2$, and $A_1A_2V_1$. In this terminology, suffixes 1 or 2 denote the first or second occurrence of the auditory or visual component of each stimulus combination. These various bimodal and unimodal stimuli were included to ensure that subjects were responding veridically on the basis of the number of perceived flashes (one or two) and not on the basis of the number of sounds. Finally, blank or no-stimulus (no-stim) trials ERPs were recorded over the same epochs as for

actual stimuli but with no stimulus presented. The reason for including the blank trials is detailed in the ERP recordings section.

The timing of the A and V components for each of the nine stimulus combinations is shown in Fig. 1B. The SOA between the two stimuli in the A_1A_2 and V_1V_2 pairs was 70 ms in every stimulus combination that included them. The A_1V_1 SOA was 10 ms in all bimodal stimulus combinations except for $A_1A_2V_1$, where V_1 followed A_1 by 200 ms. This $A_1A_2V_1$ stimulus with the delayed flash did not produce an illusory second flash and thus served as a stimulus-matched behavioral control for the $A_1V_1A_2$ test stimulus that did produce the illusion, thereby ensuring that reports of the visual illusion were not based on simply counting the number of sounds.

Stimuli were presented in 16 blocks with 20 trials of each of the nine stimulus combinations occurring on each block in a randomized sequence. All stimuli occurred with equal probability and were presented at irregular intervals of 1200-1800 ms. Subjects were instructed to report the number of flashes perceived (one or two) after each stimulus combination that contained one or more flashes. No responses were required to the unimodal auditory stimulation.

Electrophysiological (ERP) Recordings

The EEG was recorded from 62 electrode sites using a modified 10-10 system montage (Teder-Sälejärvi et al., 2005). Horizontal and vertical electro-oculograms (EOGs) were recorded by means of electrodes at the left and right external canthi and an electrode below the left eye, respectively. All electrodes were referenced to the right mastoid electrode. Electrode impedances were kept below 5 kΩ.

All signals were amplified with a gain of 10,000 and a bandpass of 0.1-80 Hz (-12 dB/octave; 3dB attenuation) and were digitized at 250 Hz. Automated artifact rejection was performed prior to averaging to discard trials with eye movements, blinks or amplifier blocking. Signals were averaged in 500 ms epochs with a 100 ms pre-stimulus interval. The averages were digitally low-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 46 Hz) to remove high frequency noise produced by muscle movements and external electrical sources. The filtered averages were digitally re-referenced to the average of the left and right mastoids.

The three-dimensional coordinates of each electrode and of three fiducial landmarks (the left and right pre-auricular points and the nasion) were determined by means of a Polhemus spatial digitizer (*Polhemus Corp., Colchester, VT*). The mean cartesian coordinates for each site were averaged across all subjects and used for topographic mapping and source localization procedures.

Neural activity associated with perception of the illusory or real second flash was isolated by subtracting the ERPs elicited by the individual unimodal components of each configuration from the ERP elicited by the total configuration, as follows:

1) Neural activity to illusory second flash: $\text{Ill\_Diff} = [(A_1V_1A_2) + \text{no-stim}] - [A_1A_2 + V_1]$

2) Neural activity to real second flash: $\text{Vis\_Diff} = [V_1V_2] - V_1$

Cross-modal interactions were also calculated for the A_1V_1 and $A_1V_1A_2V_2$ configurations, as follows:

1) A_1V_1_Diff = [(A_1V_1 + no-stim)] - [A_1 +V_1]

2) $A_1V_1A_2V_2$_Diff = [($A_1V_1A_2V_2$ + no-stim)] - [A_1A_2 +V_1V_2]

The blank or no-stimulus ERP (no-stim) was included in the calculation of these cross-modal difference waves to balance any prestimulus activity (such as a negative going anticipatory CNV that may extend into the post-stimulus period) that was present on all trials. If the no-stim trials were not included such activity would be added once but subtracted twice in the difference wave, possibly introducing an early deflection that could be mistaken for a true cross-modal interaction (Teder-Sälejärvi et al., 2002; Talsma & Woldorff 2005; Gondan & Röder 2006).

Data Analysis

ERP components observed in the Ill_Diff and Vis_Diff difference waves were first tested for significance with respect to the prestimulus baseline and compared by t-tests over all subjects (n=34). The scalp distributions and underlying neural generators of these components were then compared using methods described below. To characterize the neural correlates of perception of the cross-modal illusory flash, both between-subject and within-subject (trial-by-trial) analyses were undertaken. For the between-subject analysis, subjects were divided into two groups according to whether they reported seeing the illusion more frequently (the "SEE" group) or less frequently (the "NO-SEE" group). The groups (n=17 in each) were divided by a median split of the behavioral distribution of illusory reports (see Fig. 2). The SEE and NO-SEE groups were equivalent in age and

gender of subjects (SEE group: 8 males, 9 females, mean age 24 yrs; NO-SEE group: 8 males, 9 females, mean age 23.8 yrs). The ERP components in the Ill_Diff difference wave for the SEE and NO-SEE groups were statistically compared with respect to amplitude and scalp distribution. For those ERP components for which significant between-group differences were found, the strengths of their intracranial sources were also subjected to statistical comparisons between the two groups (see Modeling of ERP Sources section). Finally, a trial-by-trial analysis was carried out in which ERPs and cortical oscillations were compared for trials on which the illusion was perceived ("SEE" trials) vs. not perceived ("NO-SEE" trials) (see Frequency domain Analysis section).

For all analyses difference wave components were quantified as mean amplitudes within specific latency windows around the peak for each identified positive difference (PD) or negative difference (ND) component with respect to the mean voltage of a 100 ms prestimulus baseline. Components in the Ill_Diff difference wave were measured at 100-132 ms (PD120), 160-192 ms (PD180), and 252-284 ms (ND270) and in the Vis_Diff difference wave at 144-176 ms (ND160), 188-220 ms (PD200), and 260-292 ms (ND275). Each of these components was measured as the mean voltage over a specific cluster of electrodes where its amplitude was maximal. The PD120 and ND160 components were averaged over 9 occipital electrode sites spanning the midline, PD180 amplitude was measured over fronto-central electrode clusters (8 in each hemisphere and 4 over midline), the ND270 and ND275 were measured over central electrode sites (8 in each hemisphere and 4 over midline), while the PD200 was averaged over 8 occipital electrode sites in the right hemisphere (contralateral to side of stimulation).

Scalp distributions of these ERP components in the Ill_Diff and Vis_Diff difference waves were compared after normalizing their amplitudes prior to ANOVA according to the method described by McCarthy and Wood (1985). For posteriorly distributed components (PD120 vs. ND160 and PD120 vs. PD200) comparisons were made over 18 occipital electrode sites (7 in each hemisphere and 4 over midline). For the other components (PD180 vs. PD200 and ND270 vs. ND275) comparisons were made over 38 electrodes spanning frontal, central, parietal and occipital sites (15 in each hemisphere and 8 over midline). Differences in scalp distribution were reflected in significant stimulus condition (Ill_Diff vs. Vis_Diff) by electrode interactions. The scalp topographies of the PD120, PD180 and ND270 components were also compared between the SEE and NO-SEE groups using the same methods.

Modeling of ERP Sources

Source localization was carried out to estimate the intracranial generators of each ERP component in the grand-averaged difference waves within the same time intervals as those used for statistical testing. Source locations were estimated by distributed linear inverse solutions based on a Local Auto-Regressive Average (LAURA, Grave de Peralta et al., 2001). LAURA estimates 3D current density distributions using a realistic head model with a solution space of 4024 nodes equally distributed within the gray matter of the Montreal Neurological Institute's (MNI's) average template brain. It makes no a priori assumptions regarding the number of sources or their locations and can deal with multiple simultaneously active sources (Michel et al., 2001). LAURA analyses were

implemented using CARTOOL software by Denis Brunet (http://brainmapping.unige.ch/cartool.php).

To visualize the anatomical brain regions giving rise to the different components the current source distributions estimated by LAURA were transformed into the standardized coordinate system of Talairach and Tournoux (1988) and projected onto a structural brain image supplied by MRIcro (Rorden and Brett, 2000) using AFNI (Analysis of Functional NeuroImaging: Cox, 1996) software.

A statistical comparison of the LAURA source estimations between the SEE and NO-SEE subject groups was performed for those ERP components that were found to differ significantly between the groups. First, the LAURA inverse solutions for the relevant components were computed for each subject in the SEE and NO-SEE group. These source estimations were then exported to AFNI, and a region of interest (ROI) was defined for statistical analysis over voxels that encompassed the grand-average source solution in both cerebral hemispheres. The mean source current strength averaged throughout the ROI space was then compared between the two groups by ANOVA.

Trial based Analysis

A trial-by-trial analysis of the ERPs elicited in association with the illusory second flash (in the Ill_Diff waveform) was performed by separating the $A_1V_1A_2$ trials on which subjects reported seeing two flashes (SEE trials) from trials on which only a single flash (NO-SEE trials) was seen. ERP difference waves were averaged separately for the SEE trials and NO-SEE trials and the SEE-minus-NO-SEE-trials double difference wave was generated for every subject. The main components in the SEE-minus-NO-SEE-trials

double difference wave were measured at 92-124 ms (ND110) and 124-156 ms (ND130). These components were quantified as the mean voltage over the same fronto-central electrode clusters as those used to measure PD180 in the Ill_Diff waveform (see Data Analysis section).

Frequency domain Analysis

To analyze oscillatory cortical activity on SEE and NO-SEE trials, the single trial EEG signal on each channel was convolved with Morlet wavelets in a 2 sec window centered at stimulus onset. Instantaneous power and phase were extracted at each time point over 91 frequency scales from 0.6 to 101 Hz incremented logarithmically (Lakatos et al., 2005). The square root of the power values were averaged over single trials to yield the total average spectral amplitude (in uV). The average spectral amplitude at each time point and frequency was baseline corrected by subtracting the average spectral amplitude in the -300 to -50 ms pre-stimulus interval (corrected separately for each frequency band) (Tallon-Baudry et al., 1998). The time frequency spectral amplitude map on NO-SEE trials was subtracted from the map for SEE trials to reveal differential activity between the two trial types. The phase locking index across trials was calculated by normalizing the complex wavelet decomposition on every trial by its absolute value, and averaging this quantity over all trials. This analysis was restricted to the SEE subject group due to paucity of SEE trials present in the NO-SEE group.

To test the significance of differences in spectral amplitude (and phase-locking) between SEE and NO-SEE trials running paired t-tests (two-tailed) were performed initially at each electrode, time point and frequency scale. This analysis revealed

significant differences within the 20-50 Hz frequency range over the occipital scalp. Since multiple point-wise t-tests may not be statistically independent of each other, the specific differences were further analyzed using ANOVA (Kiebel et al., 2005) within intervals spanning the observed difference maxima (108-144 ms over a cluster of 12 occipital electrode sites (6 in each hemisphere) in the 25-35 Hz frequency range and 204-236 ms over 16 occipito-parietal sites (8 in each hemisphere) in the 32-40 Hz range).

RESULTS

Behavioral results

Subjects indicated by pressing one of two buttons the number of flashes perceived (one or two) for each stimulus combination that contained one or more flashes. Mean percentages of correct responses and reaction times over all 34 subjects are given in Table 1. Subjects reported perceiving an illusory second flash on an average of 37% of the $A_1V_1A_2$ trials, in agreement with the findings of Watkins et al. (2006). In contrast, the percentage of incorrect (two-flash) responses to the $A_1A_2V_1$ control stimulus having the delayed flash was much lower (9%) ($A_1V_1A_2$ vs. $A_1A_2V_1$: $F(1,33) = 52.98$, $p< 0.0001$), even though this stimulus contained the same unimodal components as the $A_1V_1A_2$ stimulus. Similarly low error rates were observed in response to the A_1V_1 (9%) and $A_1V_1A_2V_2$ (13%) stimuli. Interestingly, for the $A_1V_1V_2$ stimulus there was also a tendency for subjects to erroneously report only seeing one flash (on 44% of the trials). This phenomenon has also been previously reported in behavioral studies (Andersen et al., 2004; Shams et al., 2005b)**.** An analysis of the ERPs associated with this "suppressed flash" effect is beyond the scope of this paper and will be reported separately.

Subjects varied considerably in the percentage of $A_1V_1A_2$ trials on which they reported seeing the illusory second flash, ranging from less than 10% to over 80% (Figure 2). In order to relate perception of the illusory flash to the various ERP measures (as described below), subjects were divided by median split into groups that reported seeing the illusion more frequently (the SEE group) and less frequently (the NO-SEE group). The SEE and NO-SEE groups naturally differed substantially in the percentage of $A_1V_1A_2$ trials on which the illusory second flash was perceived (57% vs. 18%, $t(32) = 8.12$, $p < 0.0001$), but these two groups did not differ significantly in percent correct performance for any of the other stimuli (see Table 1). Reaction times between the SEE and NO-SEE groups also did not differ for the $A_1V_1A_2$ trials ($t(32) = 1.18$, p = n.s.) or for any of the other stimuli.

ERP Results

Fig. 3A shows the grand-averaged ERPs (over all 34 subjects) elicited by the illusion-inducing $A_1V_1A_2$ stimulus and by its unimodal components, A_1A_2 and V_1. The auditory ERP to A_1A_2 showed the typical pattern of P1 (60 ms), N1 (105 ms) and P2 (180 ms) components at central electrode sites. The visual ERP to V_1 also showed characteristic P1 (120 ms), N1 (180 ms) and P2 (200 ms) components, with maxima at occipital electrode sites. Both auditory and visual evoked components could be discerned in the ERP waveform elicited by the bimodal $A_1V_1A_2$ stimulus.

The Ill_Diff difference waves associated with perception of the illusory flash (see Methods) are also shown in Fig 3A for each electrode site. The earliest significant component in these difference waves was a positivity in the 100-132 ms time interval

peaking at 120 ms after the onset of A_1. This PD120 component had a bilateral distribution over the occipital scalp (Fig. 3B). The PD120 was followed by a larger positivity in the 160-192 ms time interval peaking at 180 ms (PD180), which had an amplitude maximum at fronto-central sites with a non-significant right hemispheric preponderance. The last component characterized within the first 300 ms of the Ill_Diff difference wave was a negativity within the 252-284 ms time interval peaking at 270 ms (ND270), which was largest over centro-parietal sites bilaterally. The mean amplitudes of these components relative to baseline are shown in Table 2. Components occurring after 300 ms were not analyzed because of the likelihood that activity related to decision making and response preparation would be confounded with activity related to cross-modal interaction and perceptual processing.

ERPs elicited by single (V_1) and double-flash (V_1V_2) stimuli are shown in Fig. 4A for central and occipital electrode sites. The Vis_Diff difference wave was calculated to reflect activity specifically elicited by the second flash as modified by the presence of the first flash (see Methods). The earliest significant component in the Vis_Diff wave was a negativity with a 144-176 ms latency peaking at 160 ms after time zero (defined as 10 ms before V_1 onset), see Fig 1. This ND160 had a maximal amplitude centered at the occipital pole (Fig. 4B). The ND160 was followed by a positivity in the 188-220 ms interval with a peak at 200 ms (PD200) that was significant only over right occipital channels, contralateral to stimulus presentation, and then by a negativity at 260-292 ms with a peak at 275 ms (ND275) that was distributed contralaterally over both occipital and central sites. These three components in the Vis_Diff difference wave appeared 80 ms, 120 ms and 195 ms, respectively, after presentation of the second flash. This timing

suggests an equivalence with the well-known visual evoked components C1 (ND160), P1 (PD200) and N1 (ND275), respectively. Mean amplitudes of the Vis_Diff components are given in Table 2.

The scalp topographies of the components in the Ill_Diff and Vis_Diff waveforms were compared following normalization according to the method of McCarthy & Wood (1985). The topography of the PD120 component in the Ill_Diff waveform differed significantly from the topographies of both the ND160 (Condition x Electrode interaction: $F(17,561) = 3.02$, $p < 0.0001$) and the PD200 (Condition x Electrode interaction: $F(17,561) = 3.74$, $p < 0.0001$) in the Vis_Diff waveform. Similarly, the topography of the Ill_Diff PD180 component significantly differed from that of the Vis_Diff PD200 (Condition x Electrode interaction: $F(37,1221) = 7.47$, $p < 0.0001$). Lastly, the Ill_Diff ND270 topography was significantly different from that of the Vis_Diff ND275 (Condition x Electrode interaction: $F(37,1221) = 1.92$, $p<0.0008$). These comparisons show that the ERP configuration associated with the illusory second flash was very different from that elicited by a veridical second flash.

Between Subject Analysis

In order to identify ERP components specifically associated with perception of the illusory flash, the Ill_Diff difference waveforms calculated over all trials were compared between the SEE and NO-SEE groups of subjects (Fig. 5). The PD120 component was found to be significantly larger in the SEE than the NO-SEE group (Fig. 5B, Table 3), whereas no group differences were found for the PD180 and ND270 component amplitudes.

A between-subjects correlation analysis was performed to further examine the relationship between the ERP components in the Ill_Diff waveform and the percentage of trials on which subjects reported the double-flash illusion. A significant correlation was found for the PD120 component, with greater amplitudes associated with higher levels of reporting the illusory second flash ($r = 0.48$, $p<0.005$). No significant correlations were found between behavioral performance and the amplitudes of the PD180 ($r = 0.03$, $p =$ n.s.) or ND270 ($r = -0.17$, $p =$ n.s.) components. In an fMRI investigation of the illusory second flash, no correlation was found between a subject's visual cortex activity and perception of the illusion (Watkins et al., 2006). This suggests that the neural activity giving rise to the PD120 may be too small and narrowly focused in time to give rise to a specific hemodynamic counterpart.

To examine whether the PD120 component was associated with other patterns of cross-modal interaction in the SEE group, difference waves were also calculated for the $A_1V_1A_2V_2$ and A_1V_1 cross-modal stimuli, referred as A_2V_2_Diff and A_1V_1_Diff respectively (see Methods). A between group comparison of the Ill_Diff and A_2V_2_Diff waveforms over occipital site Oz is shown in Fig. 6. For the SEE group PD120 (measured over 100-132 ms) was significant with respect to baseline in both the Ill_Diff ($t(16) = 4.18$, $p < .0008$) and the A_2V_2_Diff ($t(16) = 3.08$, $p < 0.008$) difference waves and was marginally significant in the A_1V_1_Diff difference wave ($t(16) = 1.81$, $p < 0.09$). For the NO-SEE group the PD120 measure did not reach significance in any of these difference waves (Fig. 6B). The scalp topographies of the PD120 in the SEE group showed similar occipital maxima in both the Ill_Diff and A_2V_2_Diff waveforms (Condition x Electrode interaction: $F(17,272) = 0.72$, $p =$ n.s.) (Fig. 6C). Beyond 150 ms,

both the A_2V_2_Diff and A_1V_1_Diff cross-modal waveforms elicited the PD180 and ND270 components in common with the Ill_Diff waveform (Table 2), and these two components did not differ in amplitude between the SEE and NO-SEE groups.

In sum, the PD120 component not only differentiated the SEE subject group from the NO-SEE group with respect to the illusion-producing $A_1V_1A_2$ stimulus, but also for other cross-modal stimulus combinations ($A_1V_1A_2V_2$ and, marginally, A_1V_1). Such a physiological difference generalizing over different stimuli may reflect inherent differences between the two subject groups in cross-modal connectivity between auditory and visual cortices that gives rise to the sound-induced visual illusion.

In contrast, there were no significant differences between the SEE and NO-SEE groups in any of the later difference wave components of the Ill_Diff waveform or in any component of the Vis_Diff waveform. Significant group differences were also not found in any of the components of the unisensory auditory (A_1) and visual (V_1V_2, V_1) ERPs that were used to calculate the interaction difference waves.

Source Analysis

The neural generators of the identified components in the Ill_Diff and Vis_Diff difference waveforms were modeled using a distributed minimum-norm linear inverse solution approach (LAURA, Grave de Peralta et al., 2001). All components were modeled using the ERPs averaged over all subjects except for the PD120, which was modeled from the ERPs averaged over the SEE group alone, as it was not detectable in the NO-SEE group. The generator sites estimated by LAURA were transformed into the standardized coordinate system of Talairach and Tournoux (1988) and superimposed on

the rendered cortical surface of a single individual's brain (Fig. 7). Talairach coordinates of the maxima of the current source distributions for each component are listed in Table 4.

The earliest component in the Ill_Diff waveform, the PD120, could be accounted for by current sources distributed bilaterally in lateral extrastriate cortex, including Brodmann's Area (BA) 18 and 19. The PD180 component showed two distinct sources of activity localizing to the region of the superior temporal gyrus bilaterally and to the right inferior frontal gyrus. The estimated sources for the ND270 component also showed activity focused bilaterally in the superior temporal gyrus (Fig. 7A).

In the Vis_Diff waveform the ND160 could be accounted for by estimated bilateral sources in BA18 on the lingual gyrus and extending dorsally to BA17 in the calcarine fissure. The source for the PD200 component was localized to the right middle occipital gyrus (BA19/37, contralateral to the side of visual stimulation). Finally, the ND275 component was accounted for by two regions of source activity, posteriorly in the fusiform gyrus (BA18, stronger in the right hemisphere), and more anteriorly in the inferior parietal lobule (Fig. 7B).

Since the PD120 was the main ERP component differentiating the SEE and NO-SEE subject groups (see Between Subject Analysis section), an additional statistical analysis was carried out to compare the PD120 current source densities between the two groups. A 28.8 cm^3 ROI was constructed that encompassed the grand-averaged current source maxima of the PD120 component (as modeled above for the SEE group) in both hemispheres. Within this ROI the mean current source density was greater for the SEE than the NO-SEE group ($F(1,32) = 5.02$, $p < 0.03$). The Talairach coordinates of the

maximal difference were ±45, -78, -14, which were in close proximity to the SEE group-averaged PD120 solution maxima as listed in Table 4.

Trial based Analysis

To further explore the relationship between ERP components and perception of the illusory second flash, a trial-by-trial analysis of the Ill_Diff waveform was carried out, wherein trials on which two flashes were reported (SEE trials) were separated from those on which only one flash was seen (NO-SEE trials). On average 304 total trials (s.e.m. 2.8) from each subject were included in the analysis. Within the SEE subject group, both SEE and NO-SEE trials were well represented, averaging 60% (183 trials) and 40% (121 trials) of the total trials respectively. However, within the NO-SEE group, there were far fewer SEE than NO-SEE trials, averaging 18% (55) and 82% (249), respectively. Given the markedly different distribution of trials between the two groups SEE vs. NO-SEE trial comparisons were made separately for the SEE and NO-SEE subject groups.

A comparison of the Ill_Diff waveforms from SEE and NO-SEE trials revealed significant differences at 92-124 ms and 124-156 ms post-stimulus onset in the SEE subject group (SEE vs. NO-SEE trials: (92-124 ms): $F(1,16) = 4.69$, $p< 0.05$; (124-156 ms): $F(1,16) = 8.25$, $p< 0.012$), but not in the NO-SEE group (SEE vs. NO-SEE trials: (92-124 ms): $F(1,16) = 3.67$, p= n.s.; (124-156 ms): $F(1,16) = 0.17$, p= n.s.). These differences between the groups were also evident in an overall ANOVA (with subject group as factor) as a group x trial type interaction (92-124 ms: $F(1,32) = 8.02$, $p< 0.008$;

124-156 ms: $F(1,32) = 4.93$, $p < 0.04$). The lack of effect in the NO-SEE group may have resulted from poor signal quality due to the low number of SEE trials.

The significant trial based waveform differences for the SEE group could be seen in the "double" difference wave obtained by subtracting the Ill_Diff waveform on the NO-SEE trials from the SEE trials (Fig. 8A), as negative components peaking at 110 ms (ND110) and 130 ms (ND130) respectively (significance with respect to pre-stimulus baseline for ND110: $t(16) = 2.20$, $p < 0.043$; for ND130: $t(16) = -2.90$, $p < 0.011$). The later deflection in the SEE-NO-SEE trial double difference wave at 204-236 ms showed was non-significant. Both the ND110 and ND130 components had amplitude maxima at fronto-central sites (Fig. 8B). The right hemispheric preponderance of the ND130 component in the grand-average voltage topography map did not reach significance. The occipital recordings (O1 and O2) showed that the PD120 did not differ between SEE and NO-SEE trials.

As the timing of the ND110 component in the SEE-NO-SEE trial double difference wave corresponds with the latency of the N1 component in the unimodal auditory ERP, the scalp voltage topography of the ND110 was compared to the N1 topography in the ERP to the A_1A_2 stimulus. These scalp distributions were not found to differ significantly ($F(37,592) = 1.29$, p = n.s.), suggesting a similarity between neural origins of the ND110 component and the auditory evoked N1 that is known to have neural generators in auditory cortex (Lütkenhöner and Steinsträter, 1998).

The neural sources giving rise to the ND110 and ND130 components were estimated using LAURA and were superimposed on the rendered cortical surface of an individual brain (Fig. 9). Concordant with its topographical similarity to the auditory

evoked N1, the source maximum for the ND110 was situated in the superior temporal gyrus, and its distribution included auditory cortex (BA41). The ND130 component was accounted for by source activity in a nearby region of the superior temporal gyrus (Table 4).

Frequency domain Analysis

Differences in oscillatory cortical activity between SEE and NO-SEE trials were analyzed within the SEE subject group using a Morlet wavelet decomposition in the time-frequency domain. A very early burst of enhanced spectral amplitude (or enhanced power: EP) was observed at 25-50 Hz on all channels and for all stimuli that had an auditory component; this effect could be attributed to the short latency (10-15 ms) reflex contraction of the post-auricular muscle affecting the mastoid reference electrode ("post-auricular reflex" in Fig. 10A). Over occipital sites, a robust burst of gamma power peaking at 130 ms (EP130) was observed for the SEE trials but not for the NO-SEE trials over the interval 110 to 145 ms. This selective enhancement for SEE trials was significant in the 25-35 Hz range over occipital electrodes (SEE vs. NO-SEE trials: $F(1,16) = 4.58$, $p< 0.05$; trial type x electrode: $F(5,80) = 5.53$, $p< 0.0003$). The spatial topography of EP130 is shown in Fig. 10B. For individual channels at which the EP130 was maximal, the SEE vs. NO-SEE effect was significant: at O2 ($t(16) = 2.77$, $p< 0.014$) and PO8 ($t(16) = 2.15$, $p< 0.05$) in the contralateral hemisphere, and I3 ($t(16) = 2.16$, $p< 0.047$) in the ipsilateral hemisphere. There was also significant phase locking across trials with respect to baseline during this interval for both trial types (SEE trials: $F(1,16) = 19.35$; $p< 0.0005$; NO-SEE trials: $F(1,16) = 12.01$; $p< 0.0032$). However, no phase

locking difference was found between the SEE and NO-SEE trials (SEE vs. NO-SEE trials: $F(1,16) = 0.88$, p= n.s.), which indicates that the EP130 effect is due to an actual increase in gamma amplitude on the SEE trials.

A significant enhancement in power at 204-236 ms (termed EP220) was also observed in the 32-40 Hz gamma range in the SEE-NO-SEE trial difference time-frequency representation ($F(1,16) = 4.56$, $p< 0.05$). Over individual occipito-parietal electrodes, EP220 was found to be significant only over contralateral sites: P4 ($t(16) = 2.43$, $p< 0.03$), P6 ($t(16) = 2.23$, $p< 0.041$), PO4 ($t(16) = 2.17$, $p< 0.046$), PO8 ($t(16) = 2.19$, $p< 0.046$), and O2 ($t(16) = 2.35$, $p< 0.033$). There was no significant phase locking found across trials for either SEE or NO-SEE trials during this interval.

The apparent differences in the grand-average SEE-NO-SEE trial difference time-frequency plots in the beta frequency range at 19-25 Hz (Fig. 10A) were not found to be statistically significant at any electrode site. Finally, it should be noted that no effect of trial type ($F(1,16) = 1.92$, p = n.s.) was found in the spectral amplitude within the -300 to -50 ms interval that was used as pre-stimulus baseline correction for all statistical analyses.

DISCUSSION

In the present experiment subjects reported seeing an illusory second flash on an average of 37% of the $A_1V_1A_2$ trials, but with a wide inter-individual variability ranging from 3% to 86%. To study the neural basis of the double flash illusion, a difference ERP was constructed to isolate the cross-modal interaction occurring on the illusion-producing trials, as follows: $Ill\_Diff = [(A_1V_1A_2 + NoStim) – (A_1A_2 + V_1)]$. This Ill_Diff difference

wave showed several major components, most notably a positive deflection at 120 ms after A_1 onset (PD120) that was localized to visual cortex and was larger for subjects who reported seeing the illusion more frequently (SEE group). A trial-by-trial analysis that separated individual trials into those where the illusion was seen versus not seen, however, did not find any difference in the occipital PD120 component. Instead the trial based analysis revealed an enlarged negativity in the 90-150 ms range on $A_1V_1A_2$ trials when the illusion was reported (SEE trials), which was localized to auditory cortex in its early phase (ND110) and to superior temporal cortex in its later phase (ND130). The SEE trials also elicited enhanced EEG gamma power (25-40 Hz) in visual cortex during the latency ranges 110-145 ms (EP130) and 200-240 ms (EP220). These findings indicate that the cortical activity underlying the double flash illusion includes a complex pattern of cross-modal interactions involving both modality-specific and non-specific areas as summarized in Fig. 11.

Three major components were found in the Ill_Diff waveforms in the first 300 ms after stimulus onset. The first positive deflection peaking at 120 ms (PD120) had a bilateral occipital scalp topography, and source localization using LAURA identified its principal neural generator in extrastriate visual cortex, although a minor striate contribution could not entirely be ruled out. A much larger positive deflection followed at 160-190 ms (PD180), which was largest at fronto-central scalp sites and was localized to sources in or near the superior temporal and inferior frontal gyri. A subsequent negativity in the 250-280 ms range (ND270) also had a principal source estimated in superior temporal cortex, which is a well-documented site of cross-modal interaction (reviewed in Calvert 2001). Across subjects the PD120 amplitude showed a significant

positive correlation with the proportion of trials on which the illusion was seen, and it was virtually absent in subjects who reported seeing the illusion infrequently. No such correlation was found for later components, PD180 and ND270, and their presence in other cross-modal difference waves evaluated here (A_1V_1_Diff and A2V2_Diff) suggests that they reflect general aspects of cross-modal interaction unrelated to perception of an illusory extra flash (Teder-Sälejärvi et al., 2002, 2005; Molholm et al., 2002; Talsma & Woldorff 2005).

The ERP pattern elicited by the illusory second flash was found to be very different from the ERP to a real second flash evaluated as Vis_Diff = $(V_1V_2) - V_1$. Components elicited in the Vis_Diff waveform were ND160, a negative deflection localized to the region of the calcarine fissure and surrounding extrastriate visual areas; PD200, with a source in lateral occipital cortex contralateral to stimulus presentation; and ND275, which was distributed broadly over occipito-parietal as well as frontal locations. The timing and spatial topography of these difference wave components suggest that they represent the standard sequence of C1-P1-N1 components that is well-documented to be evoked by visual stimuli (Di Russo et al., 2002, 2003; Clark and Hillyard, 1996).

The ERP correlates of illusory and real flash perception have been previously compared (Shams et al., 2001) using three recording channels over occipital cortex. Both the illusory and real flash difference waveforms were reported to display a positive deflection around 200 ms, which was interpreted as a common neural basis for the illusory and real percepts. In the present study similar positive deflections at occipital sites were also observed in the Vis_Diff and Ill_Diff waveforms at around 200 ms, but these represented the near field of the PD200 and the far field of the PD180 component,

respectively. Thus, the present whole-head recordings showed that the voltage topography and underlying neural generators of these two components were quite different.

The occipitally distributed PD120 component that correlated across subjects with the proportion of illusion trials was elicited rapidly (within 30-60 ms) after the second sound in the $A_1V_1A_2$ stimulus configuration. Arden et al. (2003) reported a similar early ERP component in the $[(A_1V_1A_2) – (A_1A_2 + V_1)]$ difference wave, albeit of the opposite polarity. However, due to the absence of a behavioral task in their experiment and their use of only a single recording site, no firm conclusions could be reached about the neural origins of this early activity or its importance for illusory flash perception. The present results suggest that the PD120 component, which was localized primarily to extrastriate visual cortex, is a strong indicator how frequently an individual subject will perceive the illusion. Recent anatomical labeling studies in macaques (Falchier et al., 2002; Clavagnier et al., 2004; Rockland and Ojima, 2003) have identified projections from primary auditory (AI) or auditory association cortices to the visual cortex (areas V1 and V2), which could mediate the cross-modal interaction responsible for producing PD120. The bilateral occipital distribution of this component despite the lateralized position of the stimuli (20º in the periphery) suggests that the audio-visual connections generating this activity are diffuse rather than spatially specific. The PD120 was also present in the other cross-modal difference wave containing two sounds (A_2V_2_Diff), but only in the SEE group. This subject-specific elicitation of the PD120 across different auditory-visual stimulus combinations may well reflect individual differences in underlying cross-modal cortical connectivity. These differences are analogous to previous electrophysiological

findings of "auditory dominant" individuals who have larger interaction effects in visual cortex at early latencies than "visually dominant" individuals (Giard and Peronnet, 1999). Overall, these differences convey the highly flexible nature of cross-modal integration across individuals, which is possibly shaped by development and experience (Bavelier and Neville, 2002).

In a trial-by-trial analysis ERPs were averaged separately for $A_1V_1A_2$ trials on which subjects reported seeing two flashes (SEE trials) vs. a single flash (NO-SEE trials). In subjects who frequently reported seeing the illusion (SEE group), the Ill_Diff difference wave showed early processing differences in the 90-150 ms interval for SEE vs. NO-SEE trials. These differences consisted of an enhanced negativity on SEE trials with two phases, an early phase (ND110) that localized to auditory cortex and a late phase (ND130) that had a source in superior temporal cortex. A trial-by-trial analysis in the time-frequency domain revealed enhanced EEG gamma (25-40 Hz) power within occipital cortex during the intervals 110-145 ms (EP130) and again at 200-240 ms (EP220) that was larger on SEE than NO-SEE trials. These findings extend previous reports of oscillatory activity associated with the double flash illusion (Bhattacharya et al., 2002) and might possibly represent an electrophysiological manifestation of the increased hemodynamic response with fMRI observed in visual cortex on illusion-producing trials (Watkins et al., 2006). The illusion related activity in superior temporal area observed in the fMRI study may similarly be related to the ND130 component reported here.

Controlling for response bias in studies of the illusory flash effect has been an issue of concern (Shams et al., 2002; Andersen et al., 2004). By showing that subjects

responded accurately on randomly interleaved control trials with auditory-visual stimulus combinations that did not produce the illusion, we ensured that response bias effects did not underlie the subjects' perceptual reports. Recently, McCormick and Mamassian (2006) provided further evidence that the illusory flash effect is a sensory based phenomenon by psychophysically demonstrating that it has a measurable contrast.

In conclusion, we investigated the neural correlates of the sound induced double flash illusion discovered by Shams and colleagues (2000, 2002) using whole head ERP recordings. We obtained evidence that the illusion is generated by a complex interaction between auditory, visual and polymodal cortical areas (Fig. 11). In those individual subjects who are disposed to see the illusion more frequently, the $A_1V_1A_2$ cross-modal interaction produces an early response in their visual cortex (PD120, onsetting 30-60 ms after A_2), which is necessary but not sufficient for seeing the illusory flash. The trigger for perceiving the illusion on a trial-by-trial basis appears to be an enhanced cross-modal interaction in the auditory cortex (ND110) that onsets even earlier (20-40 ms after A_2) and is much larger on the SEE than NO-SEE trials. We propose that the illusory percept is generated as a consequence of the interplay between these early cross-modal interactions in modality specific cortical areas, which are followed almost immediately by an enhanced burst of gamma band EEG power in visual cortex (EP130) and an increased negativity in superior temporal polymodal cortex (ND130). These findings highlight the role of rapid interactions among unimodal and polymodal cortical areas in achieving multisensory integration (Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006). A challenge for the future is to determine which aspect of this complex pattern of cross-modal interaction constitutes the immediate neural correlate of the illusory percept.

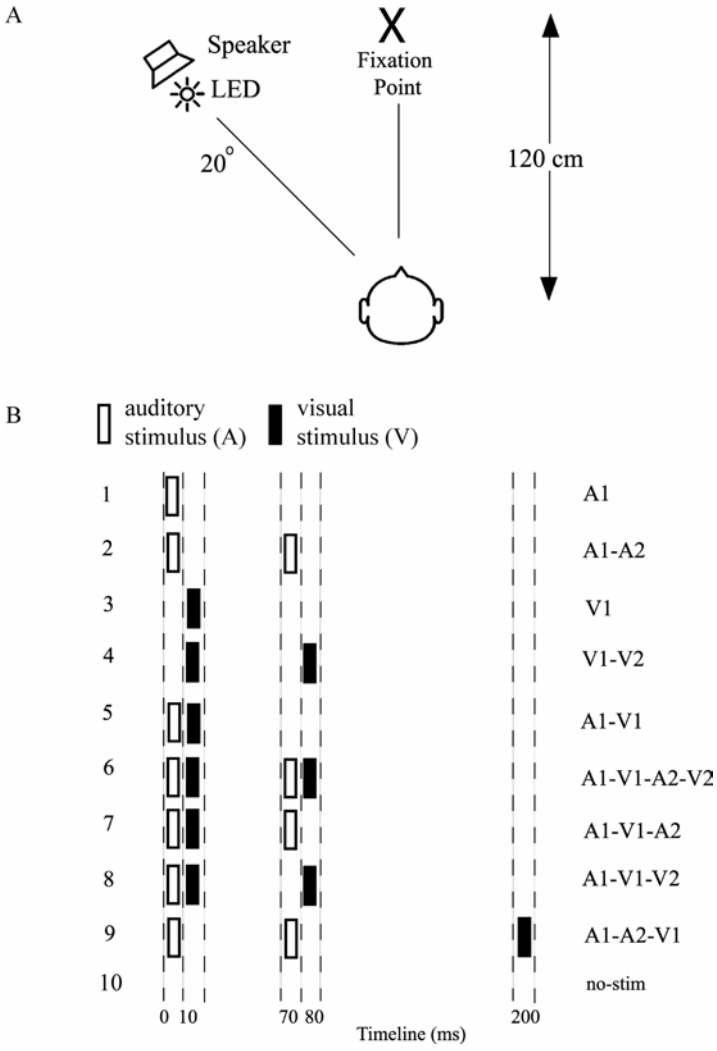

Figure 2.1 Overview of experimental design [A] Schematic diagram of experimental set-up [B] Listing of the ten different stimulus configurations, which were presented in random order. Abscissa indicates times of occurrence of auditory (open bars) and visual (solid bars) stimuli. Auditory (A) and visual (V) stimuli are labeled 1 or 2 to designate their first or second occurrence in each configuration.

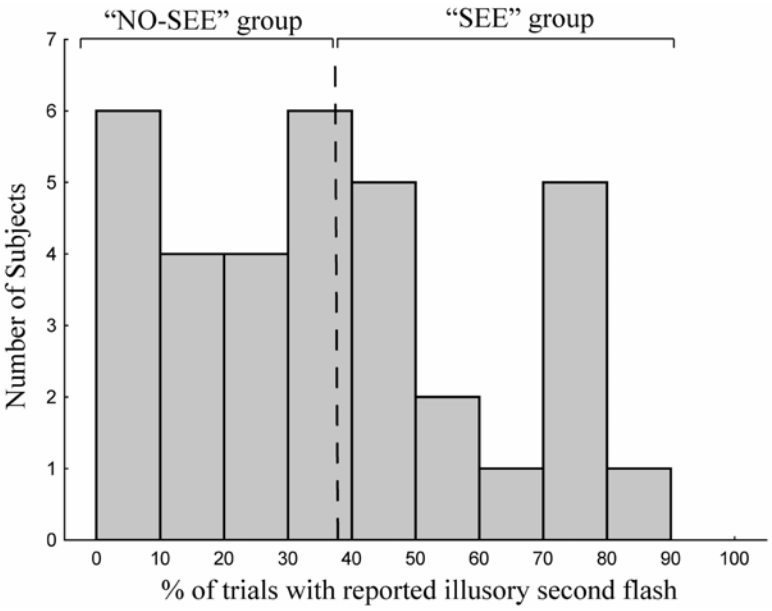

Figure 2.2 Histogram of number of subjects who reported seeing the illusory second flash to the $A_1V_1A_2$ stimulus on different percentages of trials. Subjects were divided by a median split into those who saw the illusion more frequently (SEE group) and less frequently (NO-SEE group).

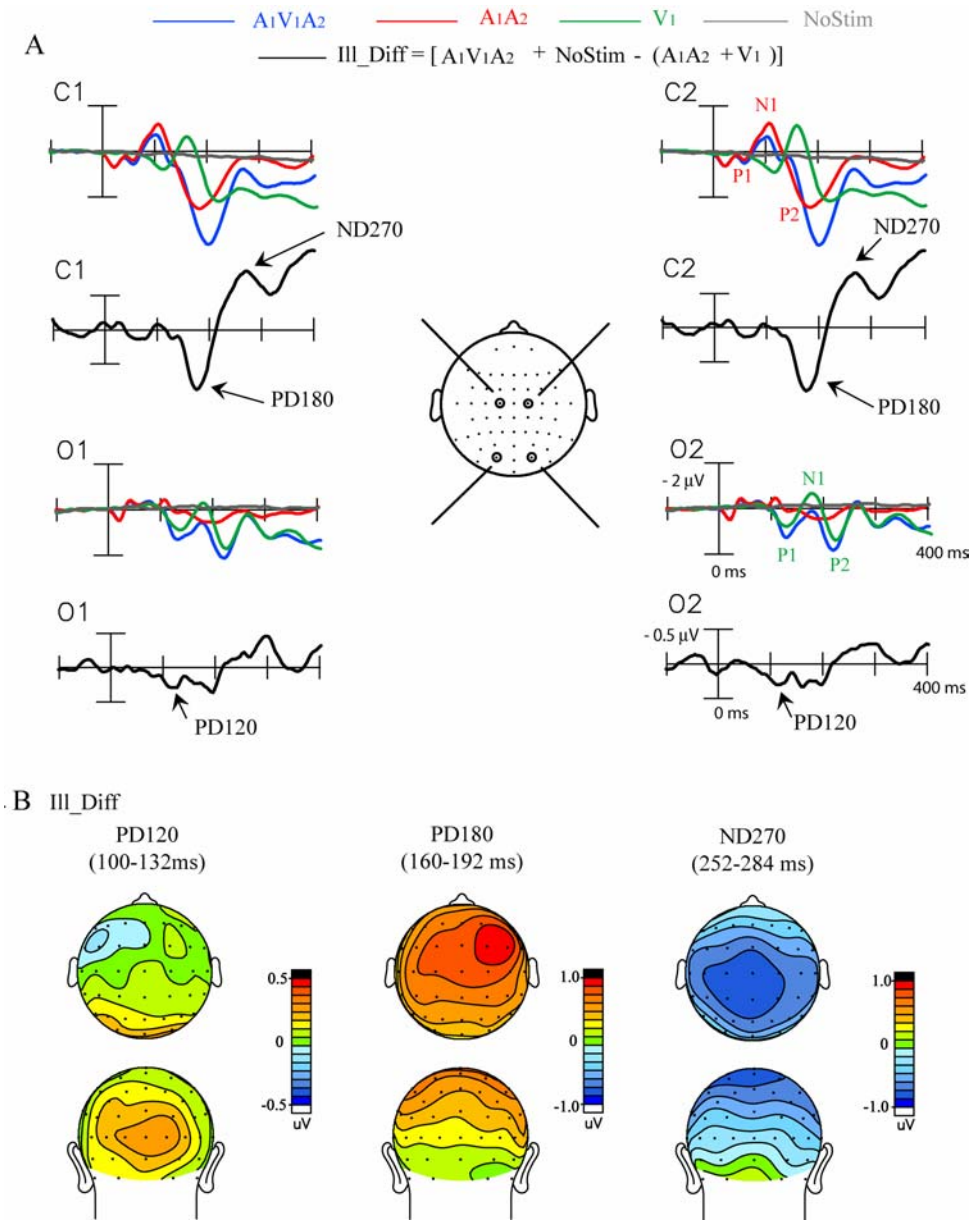

Figure 2.3 Grand-average ERPs (n=34) associated with the sound-induced illusory flash. [A] ERPs elicited by the illusion-inducing $A_1V_1A_2$ stimulus and by its unimodal constituents A_1A_2 and V_1, together with the ERP time-locked to the blank 'No-Stim' event. The Ill_Diff difference wave (see Methods) reflects the cross-modal interactions giving rise to the illusory second flash. Recordings are from left and right central (C1,2) and occipital (O1,2) sites. [B] Topographical voltage maps of the three major components in the Ill_Diff difference wave shown in top and back views.

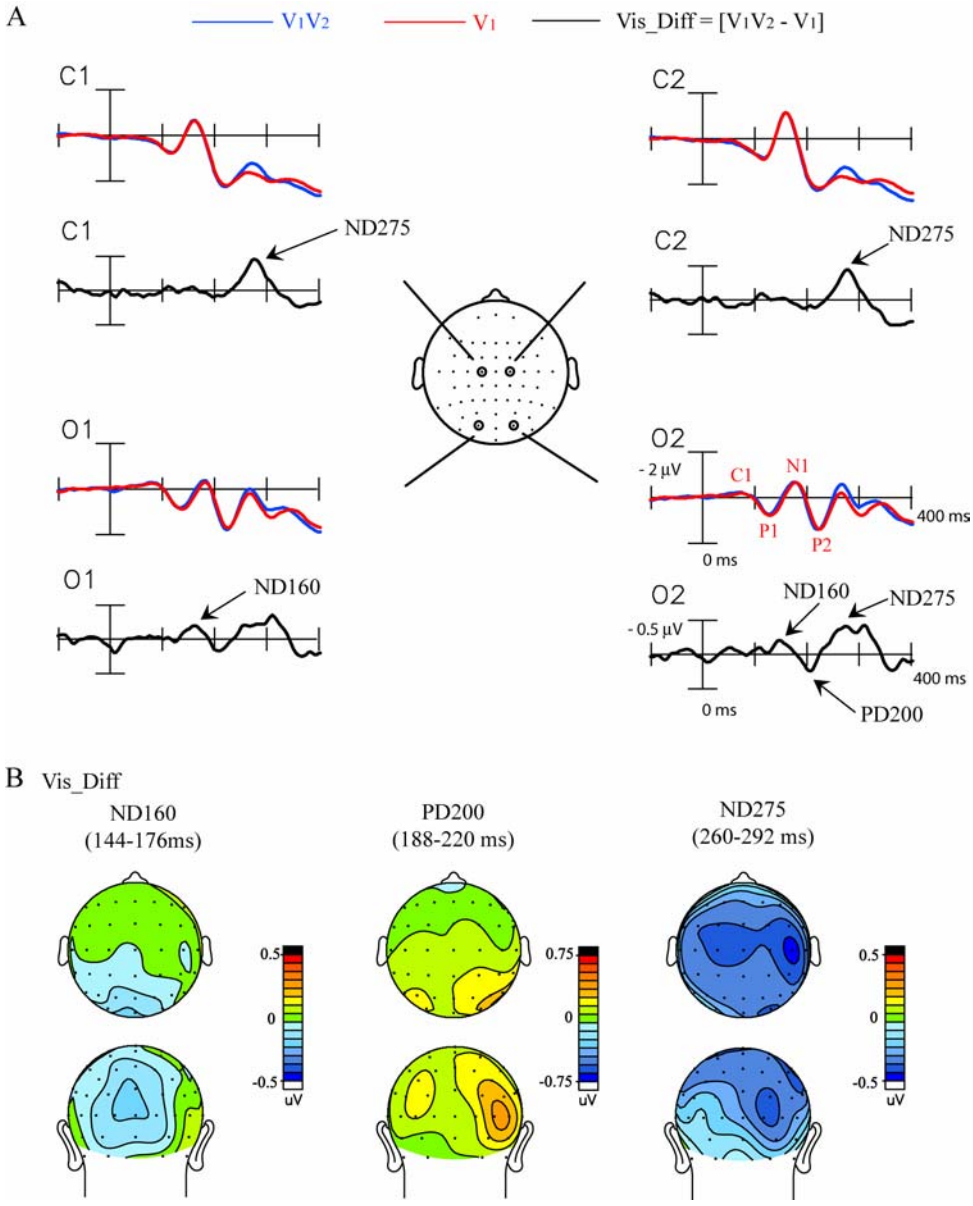

Figure 2.4 Grand-average ERPs (n=34) associated with the veridical second flash [A] ERPs elicited by the pair of flashes, V_1V_2 and by the single flash, V_1. The Vis_Diff difference wave reflects neural activity elicited by the second flash, V_2. Recordings are from left and right central (C1, 2) and occipital (O1,2) sites. [B] Topographical voltage maps of the three major components in the Vis_Diff difference wave shown in top and back views.

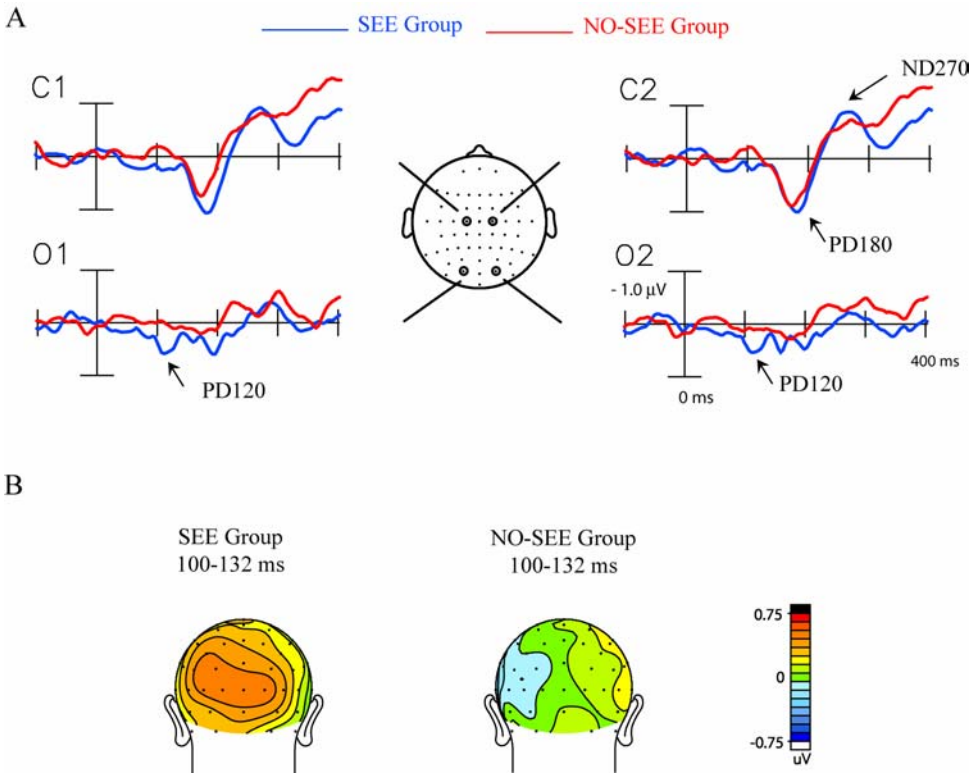

Figure 2.5 ERP differences between the SEE and NO-SEE groups. [A] Ill_Diff difference waves averaged separately for the SEE group (n=17) and the NO-SEE group (n=17). Recordings are from left and right central (C1, 2) and occipital (O1, 2) sites. [B] Voltage maps in back view comparing the topography of the PD120 component in the Ill_Diff difference waves in the two groups.

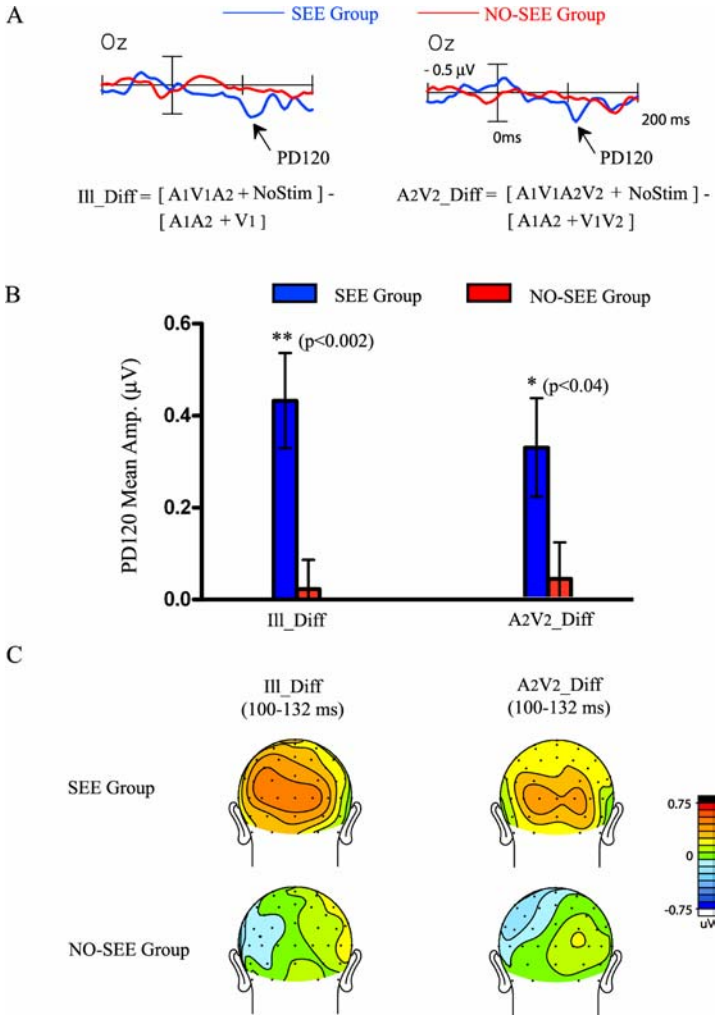

Figure 2.6 Comparison of the PD120 component elicited in the Ill_Diff and A_2V_2_Diff cross-modal difference waves for the SEE and NO-SEE groups. [A] Waveforms of the Ill_Diff and A_2V_2_Diff difference waves for the two groups recorded from an occipital electrode (Oz). [B] Bar graphs comparing the mean amplitude of PD120 in the interval 100-132 ms in the Ill_Diff and A_2V_2_Diff waveforms for the two groups. [C] Voltage maps comparing PD120 topographies for the two groups.

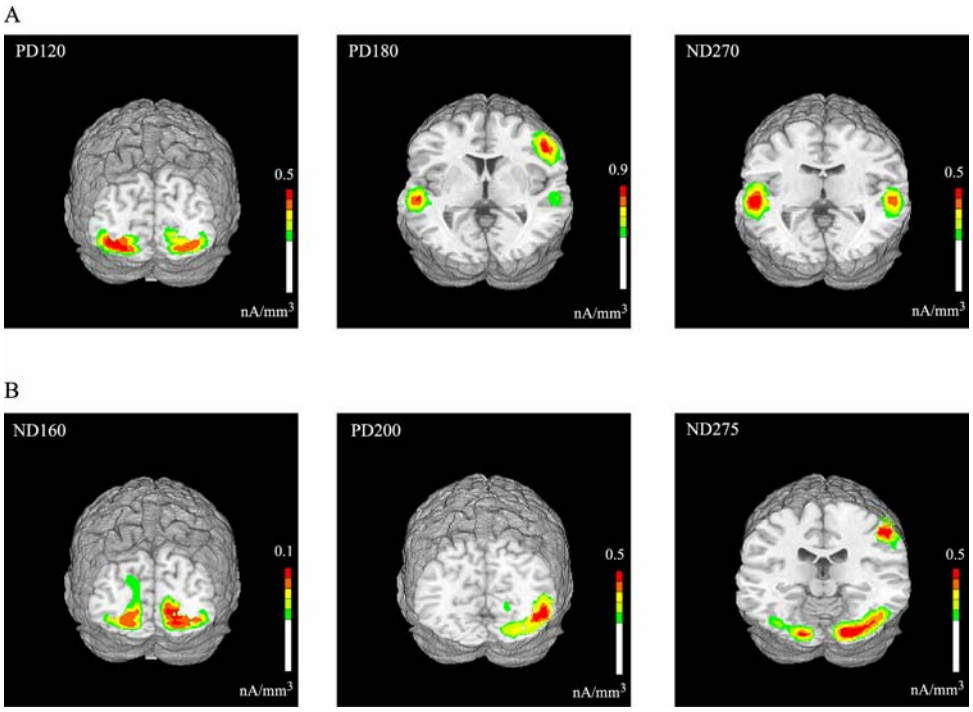

Figure 2.7 Estimated sources for the major components in the grand-average (A) Ill_Diff and (B) Vis_Diff waveforms modeled using LAURA. Results are shown on a standard fMRI rendered brain in Talairach space. LAURA inverse solutions are represented in units of current source density (nA/mm^3).

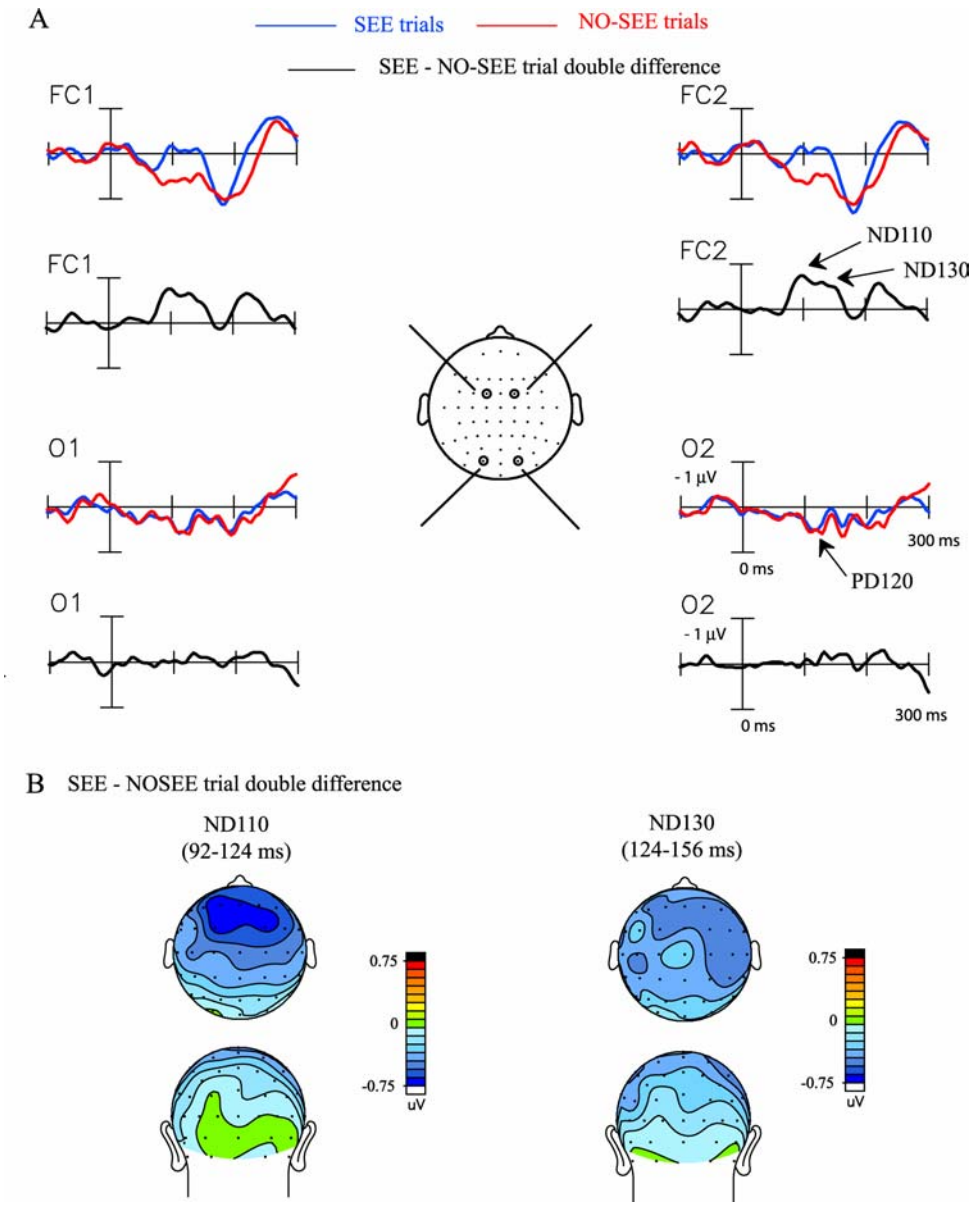

Figure 2.8 ERP differences between SEE and NO-SEE trials within the SEE subject group. [A] Ill_Diff difference waves within the SEE group averaged separately for SEE and NO-SEE trials. The SEE-NO-SEE trial double difference wave reflects differential neural activity elicited on the SEE trials with respect to NO-SEE trials. Recordings are from left and right fronto-central (FC1,2) and occipital (O1,2) sites. [B] Topographical voltage map of the two major components, ND110 and ND130 in the SEE-NO-SEE trial double difference wave shown in top and back views.

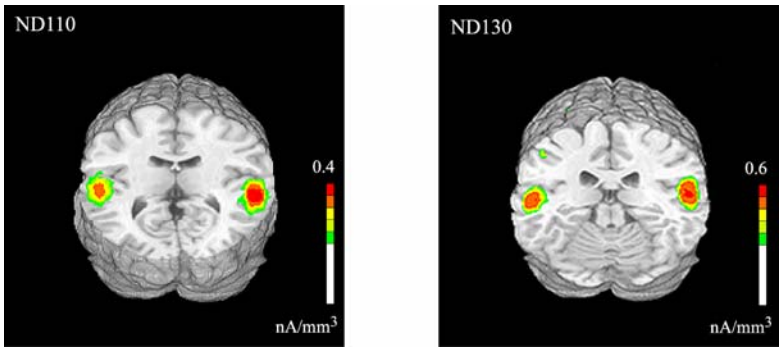

Figure 2.9 Estimated sources for the two early components in the SEE-NO-SEE trial double difference wave modeled in the SEE group using LAURA. Results are shown on a standard fMRI rendered brain in Talairach space. LAURA inverse solutions are represented in units of current source density (nA/mm^3).

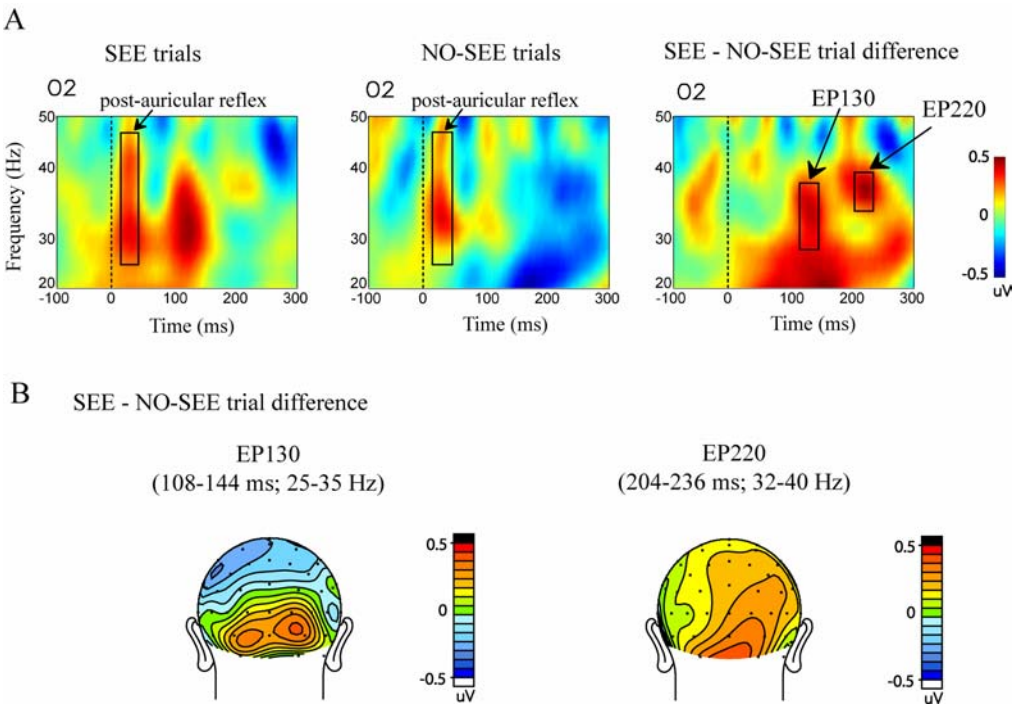

Figure 2.10 Frequency domain activity associated with perception of the sound induced illusory flash in the SEE subject group. [A] Time-frequency representation of the total average spectral amplitude on SEE trials, NO-SEE trials and the SEE-NO-SEE trial difference from an occipital electrode (O2). [B] Spatial topography maps of the two time-frequency blocks of differential spectral amplitude, EP130 and EP220 (EP: enhanced power), found in the SEE-NO-SEE trial difference shown in back view.

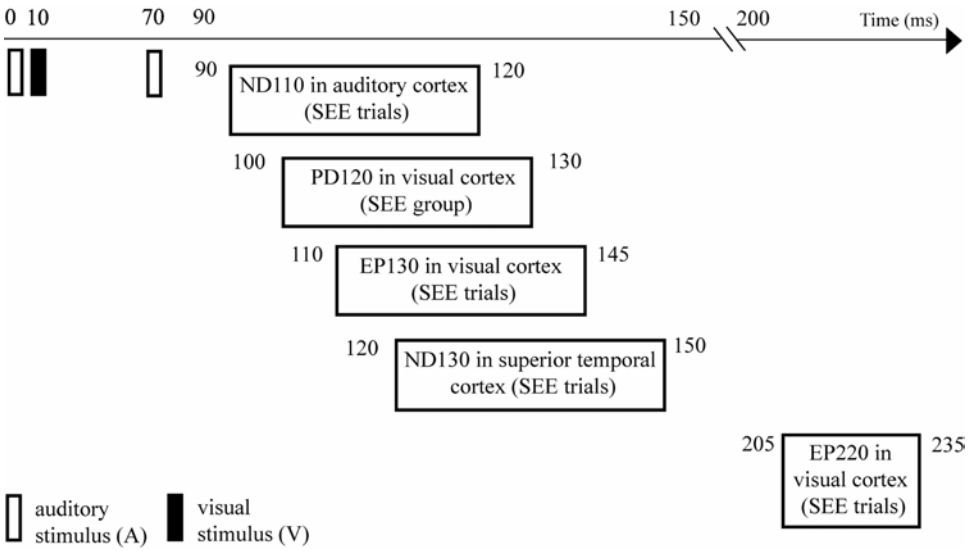

Figure 2.11 Summary of temporal progression of early cortical activity found to be associated with the sound induced extra flash illusion.

Table 2.1 Mean behavioral performance for reporting the number of flashes seen (one or two) for stimulus combinations containing one or two visual stimuli. Percent trials on which two flashes were reported and the standard error of these percentages (SEM) are reported over all 34 subjects and separately (in parentheses) for the SEE and NO-SEE subject groups (n= 17 each).

Stimulus	Percent of trials reporting two flashes over all subjects (SEE grp. / NO-SEE grp.)	SEM (% trials) over all subjects (SEE grp. / NO-SEE grp.)	Mean RT (ms) over all subjects	SEM RT (ms) over all subjects
V_1	13 (16/10)	1.9 (2.6/2.6)	612	11
V_1V_2	67 (60/74)	3.5 (4.5/5.1)	660	13
A_1V_1	9 (11/6)	1.1 (1.5/1.6)	591	14
$A_1V_1A_2V_2$	87 (86/88)	1.7 (2.7/2.0)	615	14
$A_1V_1A_2$	37 (57/18)	4.2 (4.2/2.6)	684	12
$A_1V_1V_2$	56 (46/65)	5.2 (6.9/7.2)	663	12
$A_1A_2V_1$	9 (11/7)	1.1 (1.7/1.3)	581	15

Table 2.2 Mean amplitudes of ERP components in the difference waves associated with the sound induced illusory flash (Ill_Diff); the real second flash (Vis_Diff); and other cross-modal interactions (A_2V_2_Diff and A_1V_1_Diff). Components were measured over scalp sites of maximal amplitude, as described in Methods. Significance levels of component amplitudes were tested with respect to the 100 ms pre-stimulus baseline.

	ERP Component	Amplitude (μV)	SEM (μV)	t(33)	p <
Ill_Diff	PD120 (100-132ms)	0.23	0.07	3.28	0.003
	PD180 (160-192ms)	0.74	0.16	4.76	0.0001
	ND270 (252-284ms)	-0.71	0.14	-4.98	0.0001
Vis_Diff	ND160 (144-176ms)	-0.13	0.05	-2.60	0.02
	PD200 (188-220ms)	0.27	0.08	3.55	0.002
	ND275 (260-292ms)	-0.36	0.12	-3.09	0.005
A_2V_2_Diff	PD120 (100-132ms)	0.19	0.07	2.67	0.02
	PD180 (160-192ms)	0.58	0.16	3.60	0.002
	ND270 (252-284ms)	-0.70	0.15	-4.57	0.0001
A_1V_1_Diff	PD120 (100-132ms)	0.11	0.07	1.76	n.s.
	PD180 (160-192ms)	0.57	0.13	4.26	0.0002
	ND270 (252-284ms)	-0.68	0.13	-5.30	0.0001

Table 2.3 Comparison of component amplitudes in the Ill_Diff waveform between the SEE and NO-SEE subject groups. Components were measured over scalp sites of maximal amplitude, as described in Methods.

ERP Component	Ill_Diff (SEE group)		Ill_Diff (NO-SEE group)			
	Amplitude (μV)	SEM (μV)	Amplitude (μV)	SEM (μV)	F(1,32)	p <
PD120	0.43	0.10	0.02	0.06	11.42	0.002
PD180	0.85	0.23	0.63	0.21	0.49	n.s.
ND270	-0.73	0.21	-0.69	0.20	0.02	n.s.

Table 2.4 Talairach coordinates and corresponding brain regions of the current source maxima as modeled by LAURA for the components in the Ill_Diff and Vis_Diff waveforms, and also for the components in the SEE-NO-SEE trial double difference wave.

	ERP Component	x (mm)	y (mm)	z (mm)	Brain Region
Ill_Diff	PD120	± 33	-73	-11	Lingual/ Fusiform gyri (including BA 18/19)
	PD180 (1)	± 51	-24	2	Superior Temporal gyrus
	PD180 (2)	+ 45	8	30	Inferior Frontal gyrus
	ND270	± 51	-29	3	Superior Temporal gyrus
Vis_Diff	ND160	± 13	-77	-1	Lingual gyrus (including BA 17/18)
	PD200	+ 46	-57	-6	Middle Occipital gyrus (including BA 19/37)
	ND275 (1)	+ 17	-85	-8	Fusiform gyrus (BA18)
	ND275 (2)	+ 47	-29	45	Inferior Parietal lobule
SEE-NO-SEE trial double difference	ND110	± 48	-32	10	Superior Temporal gyrus (including BA 41)
	ND130	± 51	-19	1	Superior Temporal gyrus

REFERENCES

Andersen TS, Tiippana K, Sams M (2004) Factors influencing audiovisual fission and fusion illusions. Brain Res Cogn Brain Res 21:301-308.

Arden GB, Wolf JE, Messiter C (2003) Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. Vision Res 43:2469-2478.

Baier B, Kleinschmidt A, Muller NG (2006) Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. J Neurosci 26:12260-12265.

Bavelier D, Neville HJ (2002) Cross-modal plasticity: where and how? Nat Rev Neurosci 3:443-452.

Bhattacharya J, Shams L, Shimojo S (2002) Sound-induced illusory flash perception: role of gamma band responses. Neuroreport 13:1727-1730.

Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci U S A 102:18751-18756.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110-1123.

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Paris 98:191-205.

Clark VP, Hillyard SA (1996) Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. J Cogn Neurosci 8:387-402.

Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. Cogn Affect Behav Neurosci 4:117-126.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162-173.

Di Russo F, Martinez A, Sereno MI, Pitzalis S, Hillyard SA (2002) Cortical sources of the early components of the visual evoked potential. Hum Brain Mapp 15:95-111.

Di Russo F, Martinez A, Hillyard SA (2003) Source analysis of event-related cortical activity during visuo-spatial attention. Cereb Cortex 13:486-499.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749-5759.

Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Percept Psychophys 63:719-725.

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473-490.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278-285.

Gondan M, Röder B (2006) A new method for detecting interactions between the senses in event-related potentials. Brain Res 1073-1074:389-397.

Grave de Peralta R, Gonzalez Andino S, Lantz G, Michel CM, Landis T (2001) Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. Brain Topogr 14:131-137.

Johnson JA, Zatorre RJ (2006) Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. Neuroimage 31:1673-1681.

Kiebel SJ, Tallon-Baudry C, Friston KJ (2005) Parametric analysis of oscillatory activity as measured with EEG/MEG. Hum Brain Mapp 26:170-177.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904-1911.

Lütkenhöner B, Steinsträter O (1998) High-precision neuromagnetic study of the functional organization of the human auditory cortex. Audiol Neurootol 3:191-213.

Macaluso E (2006) Multisensory processing in sensory-specific cortical areas. Neuroscientist 12:327-338.

Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli RA (2006) Multisensory Interactions within Human Primary Cortices Revealed by BOLD Dynamics. Cereb Cortex.

McCarthy G, Wood CC (1985) Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol 62:203-208.

McCormick D, Mamassian P (2006) What does the illusory flash look like? J Vision 6:189a

McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA (2003) Neural substrates of perceptual enhancement by cross-modal spatial attention. J Cogn Neurosci 15:10-19.

McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA (2005) Neural basis of auditory-induced shifts in visual time-order perception. Nat Neurosci 8:1197-1202.

Meylan RV, Murray MM (2007) Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. Neuroimage 35:244-254.

Michel CM, Thut G, Morand S, Khateb A, Pegna AJ, Grave de Peralta R, Gonzalez S, Seeck M, Landis T (2001) Electric source imaging of human brain functions. Brain Res Brain Res Rev 36:108-118.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115-128.

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex 15:963-974.

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078-1093.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19-26.

Rorden C, Brett M (2000) Stereotaxic display of brain lesions. Behav Neurol 12:191-200.

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. Curr Opin Neurobiol 15:454-458.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308.

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

Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. Neuroreport 12:3849-3852.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147-152.

Shams L, Iwaki S, Chawla A, Bhattacharya J (2005a) Early modulation of visual cortex by sound: an MEG study. Neurosci Lett 378:76-81.

Shams L, Ma WJ, Beierholm U (2005b) Sound-induced flash illusion as an optimal percept. Neuroreport 16:1923-1927.

Stein BE, London R, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci 8:497-506.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Tallon-Baudry C, Bertrand O, Peronnet F, Pernier J (1998) Induced gamma-band activity during the delay of a visual short-term memory task in humans. J Neurosci 18:4244-4254.

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098-1114.

Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex 17:679-690.

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106-114.

Teder-Sälejärvi WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. J Cogn Neurosci 17:1396-1409.

Vroomen J, de Gelder B (2004) Temporal ventriloquism: sound modulates the flash-lag effect. J Exp Psychol Hum Percept Perform 30:513-518.

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. Neuroimage. 31:1247-1256.

Chapter 3: Cortical Processes Underlying Sound-Induced Flash Fusion

ABSTRACT

When two brief flashes presented in rapid succession (< 100 ms apart) are paired with a single auditory stimulus, subjects often report perceiving only a single flash (Andersen et al., 2004, Shams et al., 2005). We used event-related potentials (ERPs) to investigate the timing and localization of the cortical processes that underlie this sound induced flash fusion, which is complementary to the sound-induced extra flash illusion that we analyzed previously (Mishra et al., 2007). The difference ERP that represented the cross-modal interaction between the visual (two flashes) and auditory (one sound) constituents of the bimodal stimulus revealed a positive component elicited 80-110 ms after the second flash, which was markedly attenuated in subjects who did not perceive the second flash. This component, previously designated as PD180 (Mishra et al., 2007), was localized by dipole modeling to polysensory superior temporal cortex. PD180 was found to covary in amplitude across subjects with the visual evoked N1 component (148-184 ms), suggesting that predisposition of individuals towards perceiving the illusion is based on differences in visual processing. A trial-by-trial analysis found that the PD180 as well as a subsequent modulation in visual cortex at 228-248 ms was diminished on trials when the two flashes were perceived as one relative to trials when flashes were correctly reported. These results suggest that the sound induced flash fusion is based on an interaction between polysensory and visual cortical areas.

INTRODUCTION

In our natural environment we constantly encounter stimulus events that have informative features in more than one sensory modality. Our sensory systems generally integrate such multimodal inputs rapidly to form a coherent percept of the sensory surroundings. The neural dynamics underlying multisensory integration have been extensively researched in electrophysiological and imaging studies, and the influence of key parameters such as spatial, temporal and semantic congruity have been characterized (Stein and Meredith, 1993; Calvert et al, 2004; Macaluso and Driver, 2005; Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006).

Interestingly, many studies have shown that our sensory systems do not always integrate external stimuli veridically. One sense may dominate another sense and influence its processing to produce perceptual illusions. For example, even though humans are generally considered to be visually dominant, there have been many reports of alteration of visual perception by audition (Stein et al., 1996; Sekuler et al., 1997; Fendrich and Corballis, 2001; Shams et al, 2000, 2002; Recanzone, 2003; Vroomen and Gelder, 2004; McDonald et al., 2003, 2005). The neurophysiological processes underlying such phenomena are only beginning to be understood. The sound-induced extra flash illusion, wherein a double flash percept results from presentation of a single flash concurrent with two rapid pulsed sounds, has been the focus of recent physiological studies (Shams et al., 2001, 2005a; Arden et al., 2003; Watkins et al., 2006; Mishra et al., 2007). In a detailed analysis of the illusion using recordings of event related potentials (ERPs) (Mishra et al., 2007) we showed that within 30-60 ms after delivery of the second sound a rapid, dynamic interplay between auditory and visual cortical areas emerged,

closely followed by activity in polymodal superior temporal cortex activity. These early cross-modal interactions predicted the subject's report of the illusory extra flash percept.

In the present study, we investigated the complement of the extra flash illusion, the so called flash fusion effect, wherein only a single flash is perceived when two brief flashes are presented in rapid succession accompanied by a single pulsed sound (Andersen et al., 2004, Shams et al., 2005b). Recently, an fMRI investigation of the flash fusion effect suggested that modulation of primary visual cortex accompanied the altered visual percept (Watkins et al., 2007). In the present study, the neural basis of sound-induced flash fusion was analyzed using 64-channel ERP recordings in conjunction with anatomical source localization. We studied the spatio-temporal patterns of neural activity associated with the flash fusion percept at two levels: between subjects who perceived the illusion and those who did not, and also within subjects on a trial-by-trial level. With the high temporal resolution of ERP recordings it was possible to investigate whether visual cortex modulation, if involved as suggested by the fMRI findings, occurs at an early input stage or via delayed feedback. These data were obtained as part of a broader ERP study that investigated the flash fusion effect, the extra flash illusion and other cross-modal interactions within the same design (Mishra et al., 2007), thereby allowing comparisons of the neural correlates of different types of intersensory interactions.

MATERIALS AND METHODS

Subjects

This paper reports additional analyses of the data obtained in the experiment previously reported by Mishra et al. (2007). Whereas our initial study was focused on the extra flash illusion, the present report analyzes the flash fusion effect observed in the same experiment. Thirty-four right-handed healthy adults (18 females, mean age 23.9 yrs) participated in the study after giving written informed consent as approved by the University of California, San Diego Human Research Protections Program. Each participant had normal or corrected-to-normal vision and normal hearing.

Stimuli and Task

The experiment, previously described in Mishra et al. (2007), was conducted in a sound-attenuated chamber having a background sound level of 32 dB and a background luminance of 2 cd/m^2. Subjects maintained fixation on a central cross positioned at a viewing distance of 120 cm. Auditory (A) and visual (V) stimuli were delivered from a speaker and red light emitting diode (LED), respectively, both positioned 20º of visual angle to the left of fixation. Each visual stimulus was a 5 ms 75 cd/m^2 flash, and each auditory stimulus was a 10 ms 76 dB noise burst. Nine different stimulus combinations were presented in random order on each block of trials. These included unimodal auditory stimuli, occurring singly (A_1) or in pairs (A_1A_2) and unimodal visual stimuli occurring singly (V_1) or in pairs (V_1V_2). Bimodal stimulus combinations included the stimulus of interest in the current study: $A_1V_1V_2$, as well as A_1V_1, $A_1V_1A_2V_2$, $A_1V_1A_2$, and $A_1A_2V_1$. In this terminology, suffixes 1 or 2 denote the first or second occurrence of

the auditory or visual component of each stimulus combination. These various bimodal and unimodal stimuli (apart from illusory percept generating stimuli: $A_1V_1V_2$ and $A_1V_1A_2$) were included to ensure that subjects were responding veridically on the basis of the number of perceived flashes (one or two) and not on the basis of the number of sounds. Finally, on blank or no-stimulus (no-stim) trials ERPs were recorded over the same epochs as for actual stimuli but with no stimulus presented.

The timing of the A and V components for each of the nine stimulus combinations was illustrated in Mishra et al. (2007, Fig. 1). Briefly, the SOA between the two stimuli in the A_1A_2 and V_1V_2 pairs was 70 ms in every stimulus combination that included them. The A_1V_1 SOA was 10 ms in all bimodal stimulus combinations except for $A_1A_2V_1$, where V_1 followed A_1 by 200 ms; this combination served as a *delayed* flash control for the $A_1V_1A_2$ stimulus that produced the extra-flash illusion.

Stimuli were presented in 16 blocks with 20 trials of each of the nine stimulus combinations occurring on each block in a randomized sequence. All stimuli occurred with equal probability and were presented at irregular intervals of 1200-1800 ms. Subjects were instructed to report the number of flashes perceived (one or two) after each stimulus combination that contained one or more flashes. No responses were required to the unimodal auditory stimulation.

Electrophysiological (ERP) Recordings

The EEG was recorded from 62 electrode sites using a modified 10-10 system montage (Teder-Sälejärvi et al., 2005). Horizontal and vertical electro-oculograms (EOGs) were recorded by means of electrodes at the left and right external canthi and an

electrode below the left eye, respectively. All electrodes were referenced to the right mastoid electrode. Electrode impedances were kept below 5 kΩ.

All signals were amplified with a gain of 10,000 and a bandpass of 0.1-80 Hz (-12 dB/octave; 3dB attenuation) and were digitized at 250 Hz. Automated artifact rejection was performed prior to averaging to discard trials with eye movements, blinks or amplifier blocking. Signals were averaged in 500 ms epochs with a 100 ms pre-stimulus interval and digitally low-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 46 Hz). The filtered averages were digitally re-referenced to the average of the left and right mastoids.

The three-dimensional coordinates of each electrode and of three fiducial landmarks (the left and right pre-auricular points and the nasion) were determined by means of a Polhemus spatial digitizer (*Polhemus Corp., Colchester, VT*). The mean cartesian coordinates for each site were averaged across all subjects and used for topographic mapping and source localization procedures.

Neural activity associated with perception of sound-induced flash fusion was isolated by calculating the cross-modal interaction between the auditory and visual components of the $A_1V_1V_2$ stimulus; in this calculation the ERPs elicited by the individual unimodal components were subtracted from the ERP elicited by the total configuration, as follows:

Neural activity associated with sound induced flash fusion:

$\text{Fusion\_Diff} = [(A_1V_1V_2) + \text{no-stim}] - [A_1 + V_1V_2]$

The blank or no-stimulus ERP (no-stim) was included in the calculation of the cross-modal difference waves to balance any prestimulus activity (such as a negative going anticipatory CNV) that was present on all trials and may extend into the early post-stimulus period. If the no-stim trials were not included such activity would be added once but subtracted twice in the difference wave, possibly introducing an early deflection that could be mistaken for a true cross-modal interaction (Teder-Sälejärvi et al., 2002; Talsma & Woldorff 2005; Gondan & Röder 2006; Mishra et al., 2007).

Data Analysis

ERP components observed in the Fusion_Diff difference wave were first tested for significance with respect to the 100 ms prestimulus baseline and compared by t-tests over all subjects (n=34). The scalp distributions and underlying neural generators of these components were then compared using methods described below. To characterize the neural correlates of perception of the cross-modal flash fusion illusion, both between-subject and within-subject (trial-by-trial) analyses were undertaken. For the between-subject analysis, subjects were divided into two groups according to whether their visual perception of the $A_1V_1V_2$ stimulus was more frequently a single flash (i.e. they reported flash-fusion: the "SEE1" group), or they reported seeing the veridical percept of two flashes more often (the "SEE2" group). The groups (n=17 in each) were divided by a median split of the distribution of percent veridical reports. The SEE1 and SEE2 groups were equivalent in age and gender of subjects (SEE1 group: 9 females, mean age 23 yrs; SEE2 group: 9 females, mean age 24.8 yrs). The ERP components in the Fusion_Diff

difference wave for the SEE1 and SEE2 groups were statistically compared with respect to amplitude and scalp distribution.

For all analyses difference wave components were quantified as mean amplitudes within specific latency windows around the peak for each identified positive difference (PD) or negative difference (ND) component with respect to the mean voltage of a 100 ms prestimulus baseline. Components in the Fusion_Diff difference wave were measured at 160-192 ms (PD180) and 224-256 ms (ND240). Each component was measured as the mean voltage over a specific cluster of electrodes where its amplitude was maximal. PD180 amplitude was measured over fronto-central electrode clusters (8 in each hemisphere and 4 over midline) and ND240 measured over similar central electrode clusters. Another component measured was the visual N1 (148-184 ms) elicited by the two unimodal visual stimuli (V_1 and V_1V_2).

Scalp distributions of these ERP components were compared between the SEE1 and SEE2 groups after normalizing their amplitudes prior to ANOVA according to the method described by McCarthy and Wood (1985). For all components comparisons were made over 38 electrodes spanning frontal, central, parietal and occipital sites (15 in each hemisphere and 8 along the midline). Differences in scalp distribution were reflected in significant group by electrode interactions. Scalp topographies of PD180 in the Fusion_Diff waveform and the visual N1 evoked by V_1V_2 were also compared in terms of the stimulus by electrode interaction.

Modeling of ERP Sources

Source localization was carried out to estimate the intracranial generators of ERP components in the grand-averaged waves within the same time intervals as those used for statistical testing. Source locations were estimated by dipole modeling using BESA (Brain Electrical Source Analysis 2000, version 5). The BESA algorithm estimates the location and the orientation of multiple equivalent dipolar sources by calculating the scalp distribution that would be obtained for a given dipole model (forward solution) and comparing it to the actual scalp-recorded ERP distribution (Scherg, 1990). The algorithm interactively adjusts (fits) the location and orientation of the dipole sources in order to minimize the relative variance (RV) between the model and the observed spatio-temporal ERP distribution. This analysis used the three-dimensional coordinates of each electrode site as recorded by a spatial digitizer. Symmetrical pairs of dipoles were fit sequentially to the components of interest; dipole pairs were constrained to be mirror-symmetrical with respect to location but were free to vary in orientation.

To visualize the anatomical brain regions giving rise to the different components the locations of BESA source dipoles were transformed into the standardized coordinate system of Talairach and Tournoux (1988) and projected onto a structural brain image supplied by MRIcro (Rorden and Brett, 2000) using AFNI (Analysis of Functional NeuroImaging: Cox, 1996) software.

Trial based Analysis

A trial-by-trial analysis of the ERPs elicited in association with flash fusion (in the Fusion_Diff waveform) was performed by separating the $A_1V_1V_2$ trials on which

subjects correctly reported seeing two flashes (SEE2 trials) from trials on which only a single flash (SEE1 trials) was seen. ERP difference waves were averaged separately for the SEE2 trials and SEE1 trials, and the SEE2-SEE1 double difference wave was generated for every subject. The grand-averaged SEE2-SEE1 waveform was calculated for 15 subjects whose behavioral SEE1 responses to the $A_1V_1V_2$ stimulus were nearest to the overall median; in these subjects the number of SEE2 and SEE1 trials were approximately the same, 54% and 46% of the total trials, respectively, while other subjects were excluded due to inequivalent trial sums in their SEE2 and SEE1 waveforms.

The main components in the SEE2-SEE1-trials double difference wave were identified in the PD180 latency range (172-200 ms) and at 228-248 ms (ND240). PD180 differences between SEE2 and SEE1 trials were quantified as the mean voltage over the same fronto-central electrode clusters as specified above. The ND240 trial differences were measured over occipital sites (6 lateral electrodes in each hemisphere) where the differences were maximal.

RESULTS

Behavioral results

The experimental paradigm was previously described in Mishra et al. (2007). Briefly, subjects maintained fixation on a central cross while unimodal visual single (V_1) or double flash (V_1V_2), unimodal auditory single (A_1) or double sound (A_1A_2) or various combinations of bimodal stimuli (A_1V_1, $A_1V_1A_2V_2$, $A_1V_1A_2$, $A_1V_1V_2$ and $A_1A_2V_1$) were presented in random order at a left peripheral location 20º lateral to fixation. Subjects

indicated by pressing one of two buttons the number of flashes perceived (one or two) for each stimulus combination that contained one or more flashes. The mean percentages of correct responses and reaction times over all 34 subjects who participated in the study have been previously reported (Mishra et al., 2007 Table 1).

For the $A_1V_1V_2$ stimulus that was the focus of the current study, perceptual reports of seeing a single flash (i.e., of flash-fusion) occurred on 44% of trials averaged over all subjects (s.e.m. 5.2%). This proportion is in close agreement with behavioral findings in the recent fMRI study of the phenomenon where flash-fusion occurred on 42% of all trials (Watkins et al., 2007). There was considerable variation, among individuals, however, in the proportion of fusion percepts, which ranged from less than 10% to over 90% (s.d. 30.2%). Hence, in order to relate the subjects' perceptual reports with brain physiology as indexed by ERPs, the 34 subject pool was divided into two groups (17 in each) by a median split of the percent fusion responses on the $A_1V_1V_2$ stimulus. The SEE1 group was the group of subjects that reported seeing flash fusion more frequently, and the SEE2 group included those who more frequently reported a veridical two-flash percept of the $A_1V_1V_2$ stimulus. Figure 1 compares the behavioral performance of the SEE1 vs. SEE2 group over all stimuli that had a visual component. The SEE1 and SEE2 groups naturally differed substantially in the percentage of $A_1V_1V_2$ trials on which flash fusion was perceived (71% vs. 18%, $t(32) = 11.2$, $p < 0.0001$), but unexpectedly these two groups also differed significantly in percent fusion responses for the V_1V_2 stimulus (41% vs. 17%, $t(32) = 6.98$, $p < 0.0001$). The groups did not significantly differ in performance for any other stimuli, nor did they show reaction time differences on any stimulus condition. In particular the SEE1 and SEE2 groups did not

differ significantly in perceiving the extra flash illusion to the $A_1V_1A_2$ stimulus (43% vs. 31%, $t(32) = 1.42$, p = n.s.). This suggests that subjects' fusion percepts were not due to a general response bias to report the number of flashes based on the number of sounds, but instead represented true perceptual reports. A significant correlation was found across subjects between percent fusion responses to the $A_1V_1V_2$ and V_1V_2 stimuli ($r(32) = 0.79$, $p < 0.0001$), suggesting that subjects who perceived the flash fusion illusion may have a general predisposition to perceive rapid double flashes as unitary. Importantly, this predisposition was not completely responsible for the flash fusion perception of the $A_1V_1V_2$ stimulus, because the presence of the A_1 sound significantly increased the perceptual reports (SEE1 group: 41% flash fusion on V_1V_2 and 71% fusion on $A_1V_1V_2$; stimulus condition x group interaction: $F(1, 32) = 38.52$, $p < 0.0001$).

ERP Results

Fig. 2A shows the grand-averaged ERPs (over all 34 subjects) elicited by the flash fusion generating $A_1V_1V_2$ stimulus and by its unimodal components, A_1 and V_1V_2. The auditory ERP to A_1 showed the typical pattern of P1 (60 ms), N1 (100 ms) and P2 (180 ms) components at central electrode sites. The visual ERP to V_1V_2 also showed characteristic P1 (120 ms), N1 (160-180 ms) and P2 (220 ms) components. Both auditory and visual evoked components could be discerned in the ERP waveform elicited by the bimodal $A_1V_1V_2$ stimulus.

The Fusion_Diff difference waves, which represent the cross-modal interaction associated with perception of sound-induced flash fusion, are also shown in Fig 2A for each electrode site. The earliest significant component in these difference waves was a

large positivity in the 160-192 ms time interval peaking at 180 ms (PD180). PD180 had an amplitude maximum at fronto-central sites with a significant right hemispheric preponderance (hemisphere effect: $F(1,33) = 11.63$, $p < 0.002$) (Fig. 2B). The other significant component characterized within the first 300 ms of the Fusion_Diff difference wave was a negativity within the 224-256 ms time interval peaking at 240 ms (ND240), which was largest over centro-parietal sites bilaterally. The mean amplitudes of these components relative to baseline are shown in Table 1. Components occurring after 300 ms in the Fusion_Diff waves were not analyzed because of the likelihood that activity related to decision making and response preparation would be confounded with activity related to cross-modal interaction and perceptual processing.

Between Subject Analysis

In order to identify ERP components specifically associated with perception of the sound induced flash fusion, the Fusion_Diff difference waveforms calculated over all trials were compared between the SEE1 and SEE2 groups of subjects (Fig. 3). The PD180 component was found to be significantly larger in amplitude in the SEE2 vs. the SEE1 group ($F(1,32) = 7.21$, $p < 0.02$) (Fig. 3B). For the SEE1 group the PD180 average amplitude did not even reach statistical significance with respect to pre-stimulus baseline (Table 2). No between-group differences were found for the ND240 component ($F(1,32) = 0.08$, $p =$ n.s.). The scalp topographies of the components were compared between groups following normalization according to the method of McCarthy & Wood (1985). The topography of the PD180 component differed between the SEE2 and SEE1 groups (Group x Electrode interaction: $F(37, 1184) = 1.49$, $p< 0.04$), with a slightly more

posterior distribution in the SEE2 group. No group differences were found in the topography of the ND240 component ($F(37,1184) = 0.25$, p = n.s.).

A correlational analysis was performed to further examine whether individual variations in PD180 amplitude corresponded with perceptual reports of the flash fusion phenomenon. A significant negative correlation was found for the PD180 component over all subjects, with greater PD180 amplitudes associated with fewer reports of the fusion effect ($r(32) = -0.39$, $p < 0.02$). No significant correlation was found between behavioral performance and the amplitude of the ND240 component ($r(32) = 0.04$, p = n.s.).

As reported by Mishra et al. (2007) the PD180 component was also observed in the other cross-modal interaction difference waves calculated for the $A_1V_1A_2$, A_1V_1, and $A_1V_1A_2V_2$ stimuli. The amplitudes of PD180 in these difference waves did not differ between the SEE1 and SEE2 groups ($A_1V_1A_2$: $F(1,32) = 2.95$, p = n.s.; A_1V_1: $F(1,32) = 2.73$, p = n.s., $A_1V_1A_2V_2$: $F(1,32) = 3.63$, p = n.s.). Thus, the PD180 component was found to differentiate the SEE1 and SEE2 groups only for the $A_1V_1V_2$ stimulus.

In the behavioral analyses (reported above) the SEE1 group showed more flash fusion responses than the SEE2 group to the V_1V_2 stimulus as well as to the $A_1V_1V_2$ stimulus. This behavioral difference was paralleled by a group difference in the visual ERP to the V_1V_2 stimulus (Fig. 4A and Table 2). The visual evoked N1 (latency range 148-184 ms) was significantly smaller for the SEE1 group compared to the SEE2 group ($F(1,32) = 4.62$, $p < 0.04$). A similar group difference was found for the N1 evoked by the single flash (V_1) stimulus (Fig. 4C, Table 2) ($F(1,32) = 4.16$, $p < 0.05$). The scalp topographies of the N1 component in the ERPs to V_1 vs. to V_1V_2 did not differ

(Condition x Electrode interaction: $F(37,1221) = 0.17$, p = n.s.). Also, there was no difference between the SEE1 and SEE2 groups in the topography of the N1 component for either visual stimulus (Group x Electrode interaction: V_1V_2: $F(37,1184) = 0.36$, p = n.s.; V_1: $F(37,1184) = 0.15$, p = n.s.).

The relationship between the visual evoked N1 and the flash fusion effect was further indicated by a significant correlation across subjects between the amplitude of the N1 to the V_1V_2 stimulus and the PD180 amplitude in the Fusion_Diff waveform ($r(32) = -0.65$, $p<0.0001$). This correlation was the physiological counterpart of the behavioral correlation described above between percent fusion responses to $A_1V_1V_2$ and V_1V_2 stimuli. The correlation across subjects between the N1 amplitude to V_1V_2 and the percent fusion responses to the V_1V_2 stimulus, however, did not reach significance ($r(32) = 0.28$, p = n.s.).

ERPs to the auditory (A_1) stimulus, which was the other sensory component of the Fusion_Diff difference wave calculation, were also analyzed for SEE2 vs. SEE1 group differences; no differences were found in any component of the auditory ERP.

Source Analysis

The neural generators of the PD180 component in the Fusion_Diff waveform and the N1 component in the V_1V_2 ERP were modeled using dipole fitting for the SEE2 subject group wherein these components were largest. Pairs of dipoles were fit to the scalp topographies of the components using the BESA algorithm (Scherg, 1990). The location of the BESA dipoles were transformed into the standardized coordinate system of Talairach and Tournoux (1988) and superimposed on the rendered cortical surface of a

single individual's brain (Fig. 6). Talairach coordinates of the dipole pairs and an estimate of their goodness of fit as reflected by residual variance are listed in Table 3.

The PD180 component in the Fusion_Diff wave localized to the region of the superior temporal gyrus bilaterally in the vicinity of Brodmann areas (BA) 21/ 22. As seen in Fig. 5B the visual N1 component elicited by V_1V_2 had two distinct phases in the 148-184 ms time range, an early phase distributed anteriorly (148-168 ms) and a late phase distributed over occipital sites (168-188 ms). The voltage distributions in the two phases were modeled sequentially using a pair of dipoles in each phase. The source of the anterior N1 distribution localized to superior temporal area in close proximity to the PD180 dipoles in the Fusion_Diff waveform. The posterior N1 localized to ventro-lateral occipital extrastriate visual cortex near the fusiform gyrus. For both phases of the N1 stronger dipole sources emerged in the right hemisphere relative to the left.

Trial based Analysis

In order to study the neural correlates of the fusion percept more directly, a trial by trial analysis of the Fusion_Diff waves was performed. This trial based analysis was carried out for 15 subjects whose behavioral reports of fusion percepts were centered around the overall median level on $A_1V_1V_2$ trials, such that each subject's SEE2 and SEE1 trial difference waves had an approximately equal number of trials (average SEE2 vs. SEE1 trials: 54%, vs. 46%).

A comparison of the Fusion_Diff waveforms between SEE2 and SEE1 trials revealed significant trial differences within the PD180 latency (172-200 ms) (SEE1 vs. SEE2 trials: $F(1,14) = 4.64$, $p< 0.05$) (Fig. 6). A later trial difference was also obtained in

the ND240 time window (228-248 ms) found to be significant over right occipital electrodes, in the hemisphere contralateral to the stimulus (SEE1 vs. SEE2 trials: $F(1,14) = 4.69$, $p < 0.05$). To distinguish this occipital effect from the anterior ND240, it will be termed $ND240_{Occ}$. These trial specific differences were evident in the "double" difference wave obtained by subtracting the Fusion_Diff waveform on SEE1 trials from SEE2 trials (Fig. 6A), as $PD180_{Diff}$ and $ND240_{OccDiff}$. Both trial specific components were significant with respect to the pre-stimulus baseline ($PD180_{Diff}$: $t(14) = 2.18$, $p < 0.05$; $ND240_{OccDiff}$ over right hemisphere: $t(14) = -2.16$, $p < 0.05$).

The voltage topography of the $PD180_{Diff}$ component was similar to that of PD180 in the Fusion_Diff wave for the 15 subjects in the trial-by-trial analysis as confirmed by the non-significant difference in their normalized spatial topographies ($PD180_{Diff}$ vs. PD180 x Electrode interaction: $F(37,518)=1.26$, p = n.s.). The later $ND240_{OccDiff}$ component had a topography centered over right visual cortex (Fig. 6B).

The neural sources giving rise to the $PD180_{Diff}$ and $ND240_{OccDiff}$ components were estimated using dipole fitting with BESA, and the talairach coordinates of the dipole pairs and their goodness of fit are listed in Table 3. The $PD180_{Diff}$ component was fit by dipole pairs with very similar coordinates as the PD180 component in the SEE2 group's Fusion_Diff wave although the $PD180_{Diff}$ had a more bilateral topography. Consistent with its occipital topography, $ND240_{OccDiff}$ was best fit by bilateral sources in visual cortex with dipoles localizing to the lingual gyrus near BA19 with a stronger right hemisphere dipole.

DISCUSSION

In this study we analyzed the neural basis of the sound-induced flash fusion phenomenon - the complement of the more extensively investigated sound-induced extra flash illusion. On average subjects reported seeing single flashes on 44% of the $A_1V_1A_2$ trials, but there was much inter-individual variability, ranging from less than 10% to over 90%. The neural basis of flash fusion was studied using ERP recordings, and the cross-modal interaction occurring on the illusion-producing trials was isolated by subtracting unimodal ERPs from the cross-modal combination ERP as follows: Fusion_Diff = $[(A_1V_1V_2 + \text{NoStim}) - (A_1 + V_1V_2)]$. The Fusion_Diff difference wave showed two major components within the 0-300 ms post-stimulus interval, a prominent positivity at 180 ms (PD180) followed by a large negativity at 240 ms (ND240). Subjects who more frequently reported perception of flash fusion had a much diminished PD180 component. A within subject trial-by-trial analysis also showed the PD180 to be markedly reduced on trials on which the two flashes within the $A_1V_1A_2$ stimulus were perceptually fused to one (SEE1 trials) vs. trials on which they were seen veridically (SEE2 trials). Using dipole modeling, PD180 was localized to polymodal cortex within the superior temporal region. The SEE2 vs. SEE1 trial comparison further revealed a reduced negativity in visual cortex at 240 ms ($ND240_{OccDiff}$) on SEE1 trials. Thus, our results suggest that the flash fusion percept is based on cross-modal interactions in polymodal cortex starting at around 100 ms after presentation of the second flash of the $A_1V_1V_2$ stimulus, which were followed about 60 ms later by differential activity in extrastriate visual cortex. The late onset of this $ND240_{OccDiff}$ suggests that it may result from feedback from polymodal

cortex, or, alternatively, from a modulation of visual evoked activity to the second flash (V_2).

The individual differences between subjects obtained in the present study, especially with respect to behavior, can potentially explain why some previous reports failed to find the sound-induced flash fusion phenomenon (Shams et al., 2002, Meylan and Murray, 2007) while others reported it to be robustly present (Andersen et al., 2004, Shams et al., 2005b, Watkins et al., 2007). In the present study a large pool of 34 participants was sampled so that the heterogeneity between subjects could be characterized, and subjects could be divided into the SEE1 or SEE2 group based on whether they perceived sound-induced flash-fusion. Shams et al. (2005b) modeled audio-visual integration using a computational model based on Bayesian statistics and proposed that the phenomena of sound-induced extra flash perception and sound-induced flash fusion both result from optimal integration between the two modalities, which differ in information reliability. For both effects the auditory stimulus was inferred to influence the visual percept because of its greater reliability in the time domain. Here we found that optimal integration was true on the average but does not necessarily apply to every subject. This was also found to hold true for the extra flash illusion (Mishra et al., 2007). Information reliability in a sensory modality may vary from one subject to another, and this diversity in cross-modal integration might possibly be shaped by development and experience (Bavelier and Neville, 2002).

The earliest cross-modal interaction component found in the Fusion_Diff waveforms was the PD180 (160-192 ms) that was localized to superior temporal cortex. The dipolar sources for this component were in close agreement with the neural

generators for the PD180 in the cross-modal interaction waveform associated with the $A_1V_1A_2$ stimulus that was previously localized using a distributed minimum-norm approach (Mishra et al., 2007). A component closely resembling the present PD180 has been found in many previous paradigms studying cross-modal interactions (Teder-Sälejärvi et al., 2002, 2005; Molholm et al., 2002; Talsma & Woldorff 2005, Mishra et al., 2007), but this is the first report to our knowledge of its modulation by perception. In particular, variations in PD180 were not associated with the extra flash illusion, either between subjects or on a trial by trial basis (Mishra et al., 2007). This suggests that the underlying audio-visual interaction in the superior temporal region is related more to the precise timing and segmenting of visual inputs than to the generation of an illusory visual percept.

Interestingly, in the present study a strong correlation was found between the subjects' perceptual reports on the unimodal V_1V_2 stimulus and the $A_1V_1V_2$ stimulus. Subjects who more frequently mis-perceived V_1V_2 as a single flash also had a greater propensity to report sound-induced flash fusion. The single flash percept on $A_1V_1V_2$ trials was not entirely driven by the paired visual stimuli, however, as illusory fusion occurred more frequently in the presence of the A_1 sound than in its absence (V_1V_2 stimulus). Paralleling these perceptual reports, the amplitude of the evoked N1 component in the ERP to V_1V_2 was correlated with the PD180 amplitudes in the Fusion_Diff waveform. In other words, subjects who perceived sound-induced flash fusion not only had smaller PD180s in the Fusion_Diff waveforms but also smaller visual N1s on V_1V_2 trials. The visual N1 was found to have early (148-168 ms) and late (168-188 ms) phases with distinct topographies, and source localization revealed that the neural generators of the

early phase were in close proximity to the PD180 source in superior temporal cortex. This suggests that the neural basis of the flash fusion effect for both V_1V_2 and $A_1V_1V_2$ trials may involve the same cortical region, which may account for their correlation. Individual differences in unisensory processing that affected multisensory interactions have been previously noted in a few studies (Giard and Peronnet, 1999, Fort et al., 2002). In those studies subjects were categorized as either "auditory dominant" or "visually dominant" based on their superior reaction times in one modality or the other, and they were found to show differential cross-modal interaction effects in auditory/ visual sensory cortices depending on which of their modalities was behaviorally dominant.

The trial-by-trial analysis of the SEE1 vs. SEE2 trial ERPs in a group who saw flash fusion with moderate frequency also confirmed diminished PD180 amplitudes on SEE1 trials as was also found for the SEE1 group. In a later time window (228-248 ms) a trial difference was also found in visual cortex, $ND240_{OccDiff}$, that localized to ventral extrastriate areas near the fusiform gyrus. Since this modulation in visual cortex occurred after the PD180 modulation in polymodal superior temporal area, it may be a result of feedback from the polymodal area. In a recent fMRI investigation Watkins et al. (2007) reported greater BOLD (blood oxygen level dependent) activity in primary visual cortex on SEE2 vs. SEE1 trials. In the present study the enhanced occipital ERP on the SEE2 trials was localized primarily to ventral extrastriate visual cortex, but a primary cortex contribution could not be ruled out. The superior temporal resolution of the ERP recordings, however, suggests that trial specific visual cortex involvement did not occur in the initial response phase but rather was probably driven by feedback from higher polymodal areas. Connectivity analyses in a recent fMRI study of audio-visual temporal

correspondence also provided evidence for feedback from the superior temporal area to primary visual cortex (Noesselt et al., 2007b).

A recent ERP study of auditory driving of visual perception used slow audio-visual flutter and flicker rates of 3-5 Hz and found that modulation of occipital visual areas occurred as late as 500 ms after stimulus onset, subsequent to modulation at parietal and frontal recording sites (Noesselt et al., 2007a). Auditory driving has been considered an extended case of the sound-induced flash fusion/ fission phenomena, and hence the later occipital modulations found by Noesselt et al. (2007a) may correspond to the trial specific occipital modulations observed in the current study. Noesselt et al. (2007a) also suggested that the late occipital modulations in their study may be a result of feedback from higher multisensory areas. Finally, a modulation within extrastriate visual areas was also observed within a similar latency range as $ND240_{OccDiff}$ by Meylan and Murray (2007) who isolated activity to the second flash V_2 of the $A_1V_1V_2$ stimulus by subtracting ERPs to A_1V_1 from the ERPs to the cross-modal stimulus. Subjects in their study did not perceive the flash fusion illusion, however, which could be due to the smaller subject pool of 8 participants or different stimulus parameters.

In conclusion, we investigated the neural correlates of the sound-induced flash fusion illusion using whole head ERP recordings. For individuals who were prone to see the two flashes of $A_1V_1V_2$ as one, the large cross-modal interaction component PD180, onsetting 80-112 ms after V_2 and localizing to polymodal superior temporal area, was greatly diminished. These subjects also showed a significantly reduced N1 component to V_1V_2 stimuli both in its early phase (148-168 ms) that was localized to the same superior temporal region as the PD180 and in its later phase (168-188 ms) localized to extrastriate

visual cortex. These subjects also reported single flash percepts more frequently to the V_1V_2 stimulus, suggesting that the cross-modal flash fusion effect is based at least in part on individual differences in visual processing. Modulation of the PD180 component was also consistently observed in the trial-by-trial analysis of the ERPs, with an attenuated PD180 on trials on which the second flash was not perceived by subjects. These results suggest that the veridical perception of the two flashes in the V_1V_2 and $A_1V_1V_2$ stimuli depends upon a larger visual evoked response and/ or cross-modal interaction in superior temporal polymodal cortex. Activation of this polysensory area may enable accurate judgments of the timing and sequencing of visual as well as auditory stimuli.

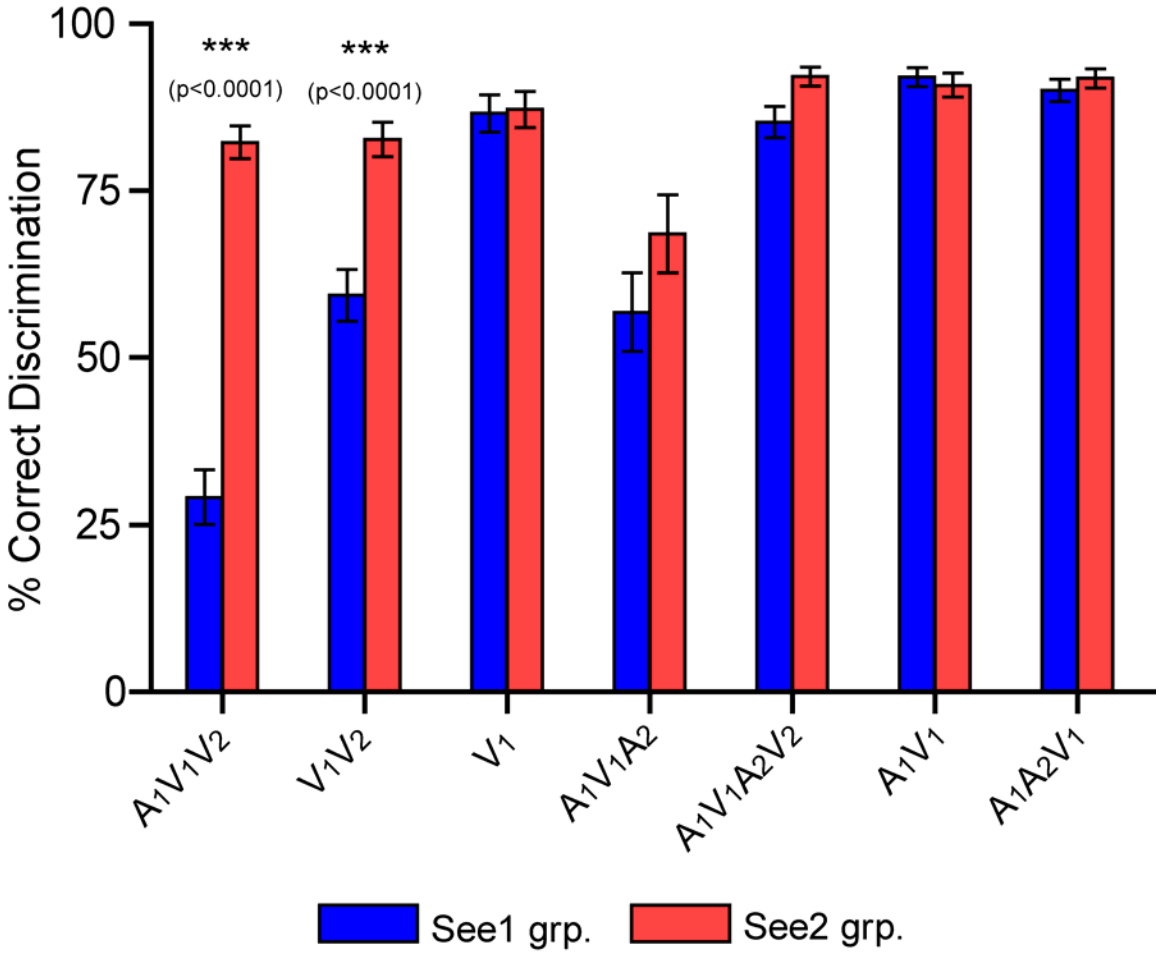

Figure 3.1 Behavioral performance comparisons across all experimental stimuli between the two groups of subjects: those who frequently perceived the two flash component of the $A_1V_1V_2$ stimulus as a single flash (SEE1 group), and those who correctly reported seeing two flashes on the majority of trials (SEE2 group).

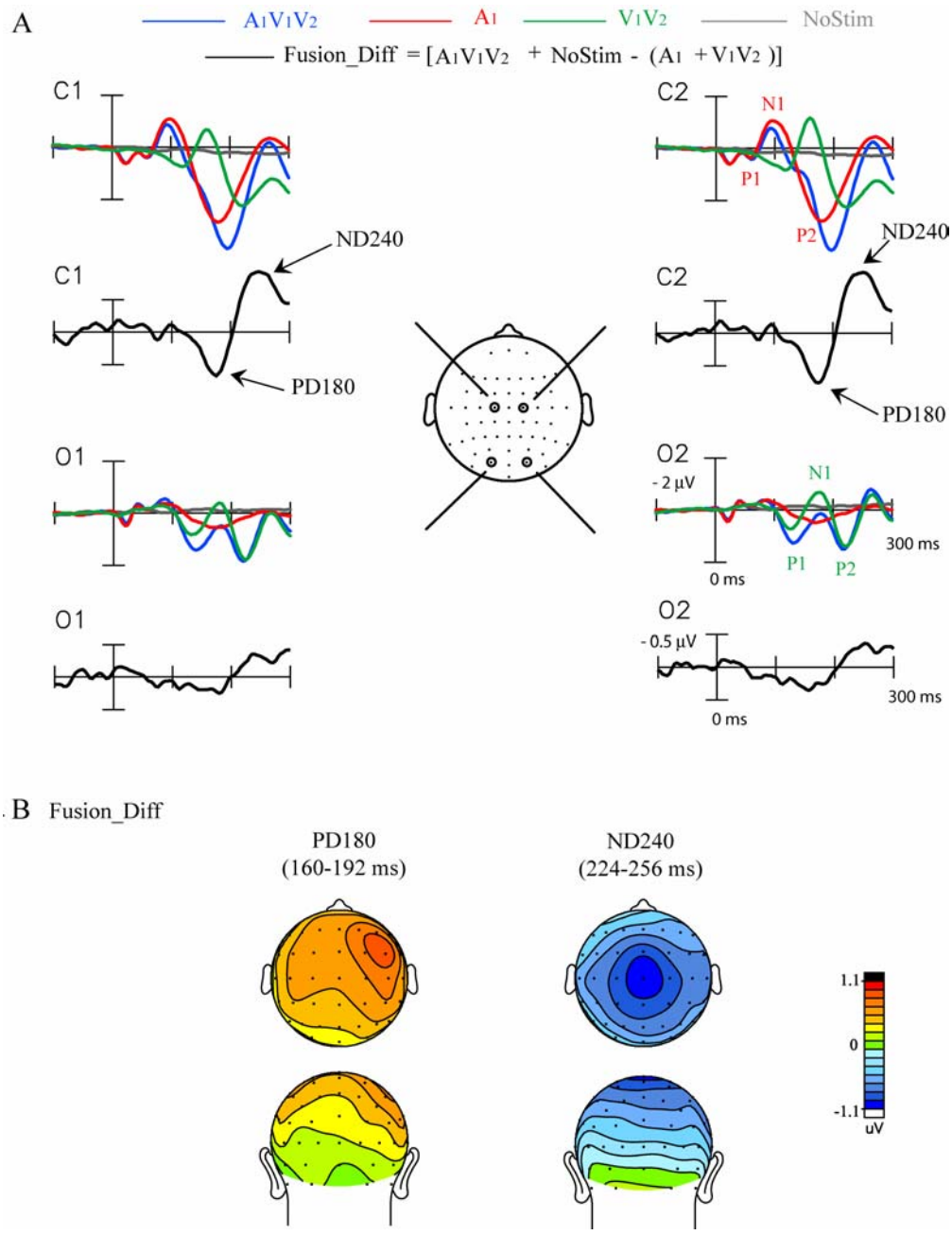

Figure 3.2 Grand-average ERPs (n=34) associated with the sound-induced flash fusion illusion. [A] ERPs elicited by the illusion-inducing $A_1V_1V_2$ stimulus and by its unimodal constituents A_1 and V_1V_2, together with the ERP time-locked to the blank 'No-Stim' event. The Fusion_Diff difference waves represent the cross-modal interactions underlying the flash fusion illusion. Recordings are from left and right central (C1,2) and occipital (O1,2) sites. [B] Topographical voltage maps of the two major components in the Fusion_Diff difference wave.

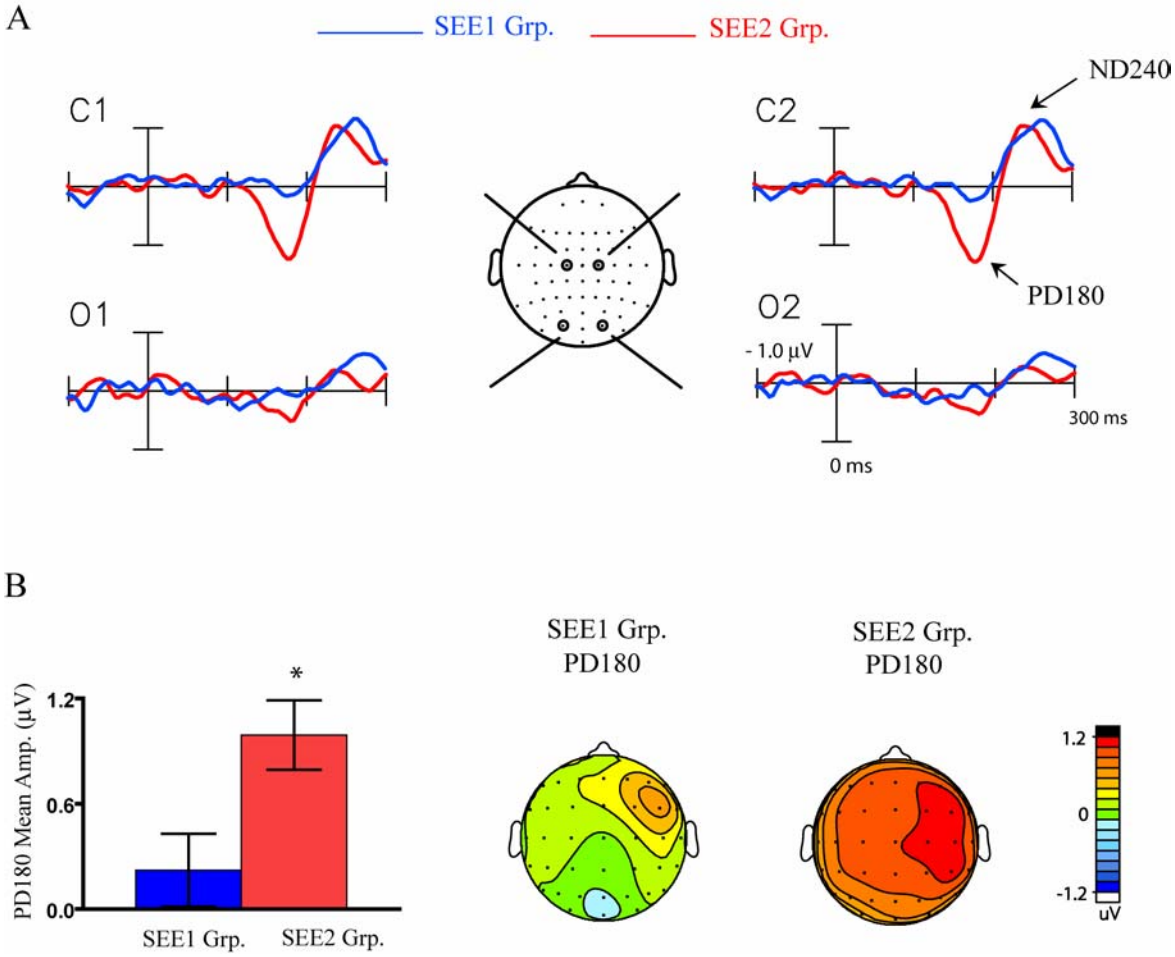

Figure 3.3 ERP differences between the SEE1 and SEE2 groups. [A] Fusion_Diff difference waves averaged separately for the SEE1 group (n=17) and the SEE2 group (n=17). Recordings are from left and right central (C1, 2) and occipital (O1, 2) sites. [B] Bar graphs comparing the mean amplitude of PD180 in the 160-192 ms interval in the Fusion_Diff waveforms for the two groups, and voltage maps showing the topography of the PD180 component in the two groups.

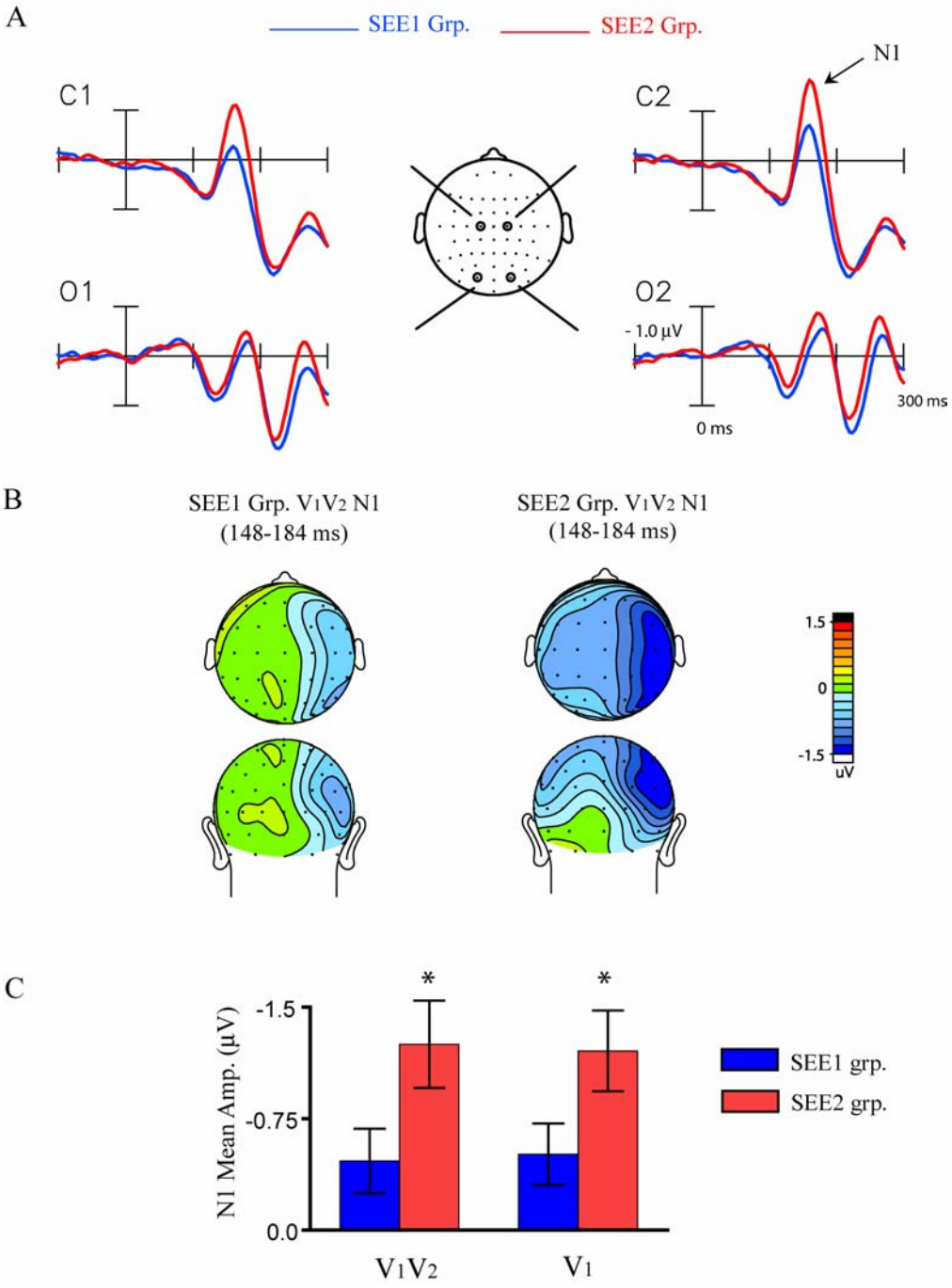

Figure 3.4 ERP differences between the SEE1 and SEE2 groups for the V_1V_2 stimulus. [A] ERPs to V_1V_2 averaged separately for the SEE1 group and the SEE2 group. Recordings are from left and right central (C1, 2) and occipital (O1, 2) sites. [B] Voltage maps comparing the topography of the visual N1 component in the two groups. [C] Bar graphs comparing the mean amplitude of the N1 component in the 148-184 ms interval between the two groups in the ERPs to the V_1V_2 and V_1 stimuli.

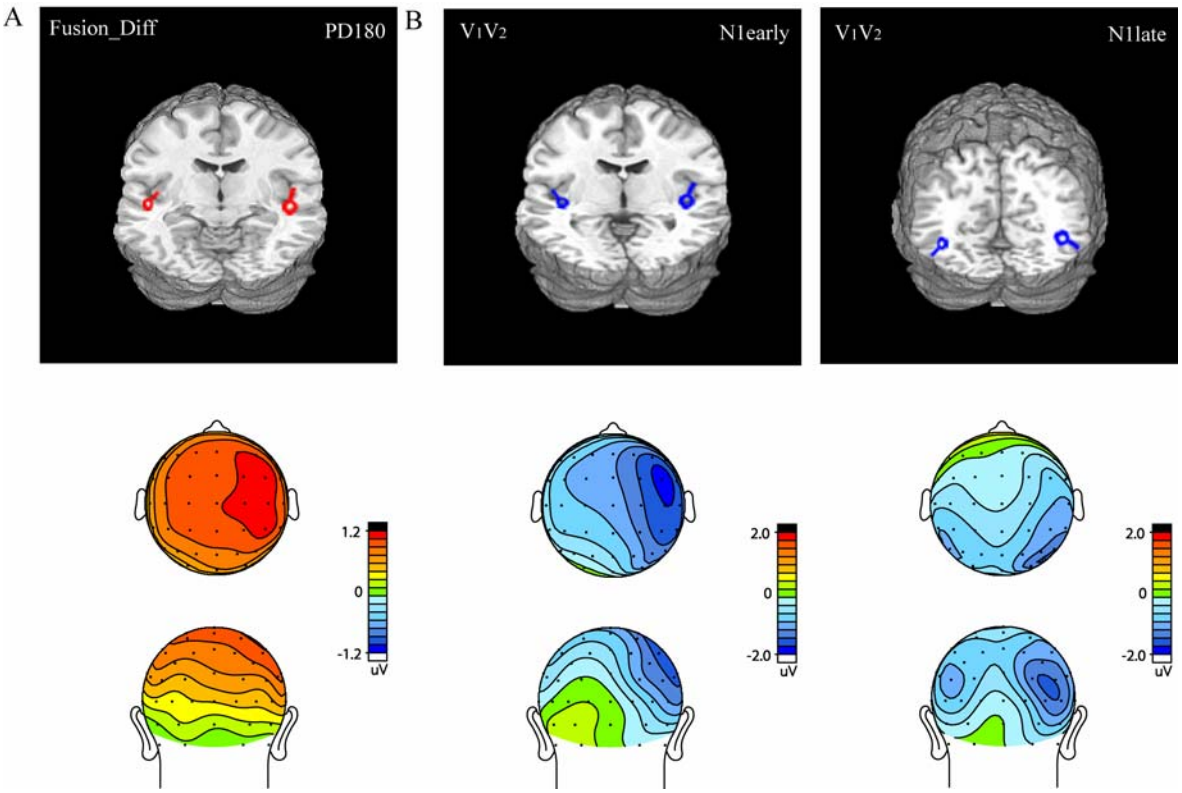

Figure 3.5 Estimated dipolar sources modeled using BESA and corresponding voltage topographies of the ERP components related to flash fusion in the SEE2 subject group. [A] Source model and topography of the PD180 component in the Fusion_Diff waveform. [B] Source models and topographies of the early (148-168 ms) and late (168-188 ms) phases of the visual N1 component evoked by the V_1V_2 stimulus. Dipole models are shown on a standard fMRI rendered brain in Talairach space.

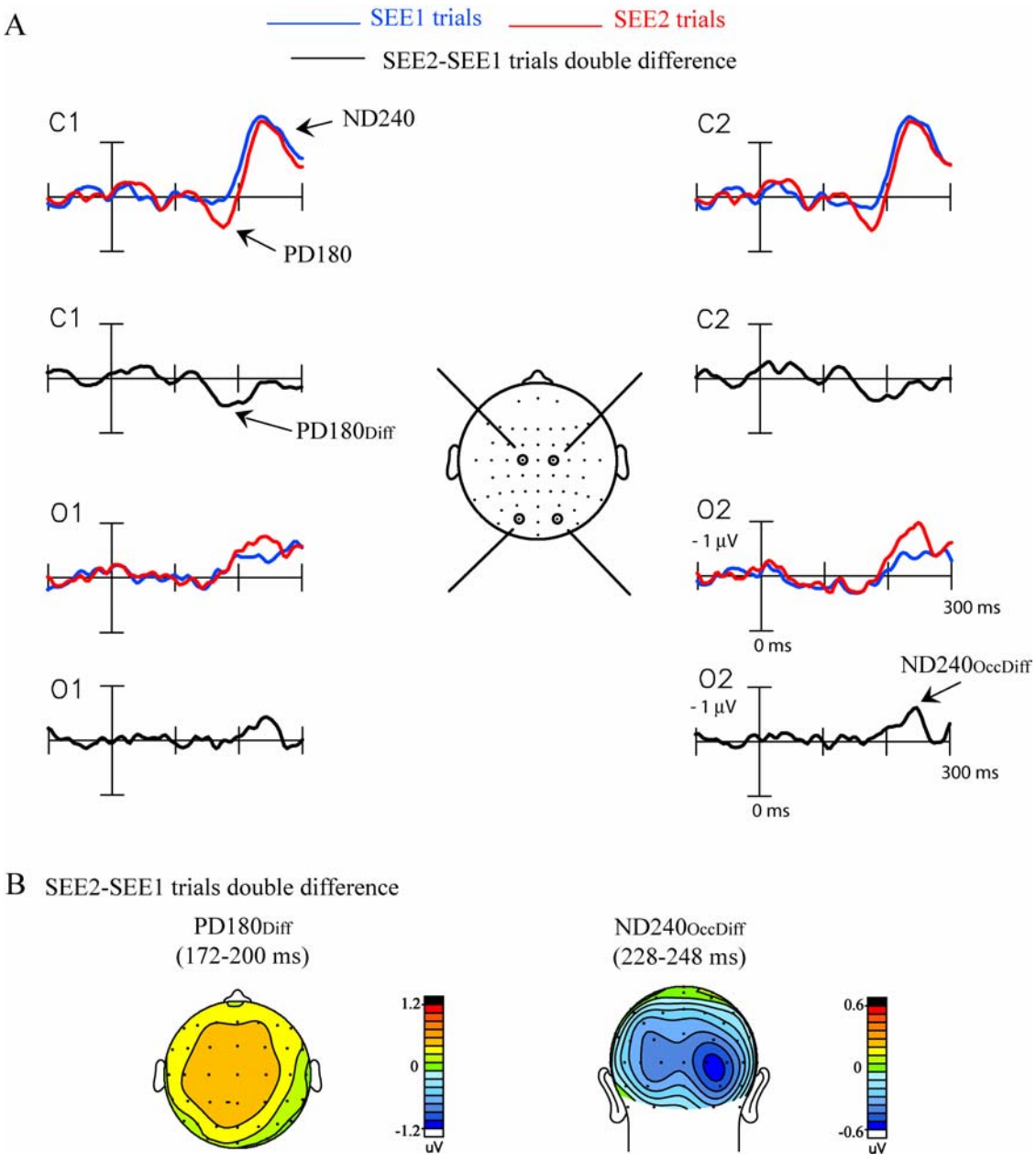

Figure 3.6 ERP differences between SEE1 and SEE2 trials for 15 subjects who had nearly equivalent numbers of SEE1/ SEE2 trials. [A] Fusion_Diff difference waves averaged separately for SEE1 and SEE2 trials. The SEE2-SEE1 trial double difference wave reflects differential neural activity elicited on the SEE2 trials vs. SEE1 trials. Recordings are from left and right central (C1,2) and occipital (O1,2) sites. [B] Topographical voltage map of the two major components, $PD180_{diff}$ and $ND240_{OccDiff}$ in the SEE2-SEE1 trial double difference wave.

Table 3.1 Mean amplitudes of ERP components in the difference waves associated with sound-induced flash fusion (Fusion_Diff) averaged over all 34 subjects. Components were measured over scalp sites of maximal amplitude. Significance levels of component amplitudes were tested with respect to the 100 ms pre-stimulus baseline.

	ERP Component	Amplitude (μV)	SEM (μV)	t(33)	p <
Fusion_Diff	PD180 (160-192ms)	0.61	0.16	3.88	0.0005
	ND240 (224-256ms)	-0.79	0.18	-4.47	0.0001

Table 3.2 Component amplitudes in the Fusion_Diff waveforms and N1 amplitudes in the visual V_1V_2 and V_1 waveforms for the SEE1 and SEE2 subject groups. Components were measured over scalp sites of maximal amplitude and tested for significance with respect to the 100 ms pre-stimulus baseline.

ERP	Comp-onent	SEE 1 grp.				SEE2 grp.			
		Amp. (μV)	SEM (μV)	t(16)	p <	Amp. (μV)	SEM (μV)	t(16)	p <
Fusion_ Diff	PD180	0.22	0.21	1.07	n.s.	0.99	0.20	5.01	0.0002
	ND240	-0.74	0.28	-2.65	0.02	-0.84	0.22	-3.73	0.002
V_1V_2	N1	-0.47	0.22	-2.15	0.05	-1.25	0.29	-4.26	0.0007
V_1	N1	-0.51	0.21	-2.47	0.03	-1.21	0.27	-4.45	0.0005

Table 3.3 Talairach coordinates and corresponding brain regions of the dipole fits as modeled by BESA for the components in the Fusion_Diff and V_1V_2 waveforms for the SEE2 subject group, and also for the components in the SEE2-SEE1 trial double difference wave. Percent residual variance not accounted for by the model over the interval specified in parentheses is shown for each component.

	ERP Component	x (mm)	y (mm)	z (mm)	Region	Res. Var. (%)
Fusion_Diff SEE 2	PD180	± 45	-11	-9	Vicinity of Superior Temporal Gyrus (STG)	3% (160-192 ms)
V_1V_2 SEE 2	N1 (148-168 ms)	± 41	-11	-1	Vicinity of STG	5% (148-188 ms)
	N1 (168-188 ms)	± 39	-62	-7	Vicinity of Fusiform Gyrus	
SEE2-SEE1 trial double difference	$PD180_{Diff}$	± 46	-6	-5	Vicinity of STG	10% (172-200 ms)
	$ND240_{OccDiff}$	± 29	-64	-1	Vicinity of Lingual gyrus	4% (228-248 ms)

REFERENCES

Andersen TS, Tiippana K, Sams M (2004) Factors influencing audiovisual fission and fusion illusions. Brain Res Cogn Brain Res 21:301-308.

Arden GB, Wolf JE, Messiter C (2003) Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. Vision Res 43:2469-2478.

Bavelier D, Neville HJ (2002) Cross-modal plasticity: where and how? Nat Rev Neurosci 3:443-452.

Calvert GA, Stein BE, Spence C (2004) The handbook of multisensory processing. Cambridge, MA: MIT.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162-173.

Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Percept Psychophys 63:719-725.

Fort A, Delpuech C, Pernier J, Giard MH (2002) Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res Cogn Brain Res 14:20-30.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278-285.

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473-490.

Gondan M, Roder B (2006) A new method for detecting interactions between the senses in event-related potentials. Brain Res 1073-1074:389-397.

Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. Trends Neurosci 28:264-271.

McCarthy G, Wood CC (1985) Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol 62:203-208.

McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2003) Neural substrates of perceptual enhancement by cross-modal spatial attention. J Cogn Neurosci 15:10-19.

McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2005) Neural basis of auditory-induced shifts in visual time-order perception. Nat Neurosci 8:1197-1202.

Meylan RV, Murray MM (2007) Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. Neuroimage 35:244-254.

Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci 27:4120-4131.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115-128.

Noesselt T, Bonath B, Boehler CN, Schoenfeld MA, Heinze HJ (2007a) On perceived synchrony-neural dynamics of audiovisual illusions and suppressions. Brain Res.

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007b) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J Neurosci 27:11431-11441.

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078-1093.

Rorden C, Brett M (2000) Stereotaxic display of brain lesions. Behav Neurol 12:191-200.

Scherg M (1990) Fundamentals of dipole source analysis. Auditory evoked magnetic fields and electric potentials (Grandori F, Hoke M, Roman GL, eds):40-69.

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. Curr Opin Neurobiol 15:454-458.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308.

Shams L, Iwaki S, Chawla A, Bhattacharya J (2005a) Early modulation of visual cortex by sound: an MEG study. Neurosci Lett 378:76-81.

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

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147-152.

Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. Neuroreport 12:3849-3852.

Shams L, Ma WJ, Beierholm U (2005b) Sound-induced flash illusion as an optimal percept. Neuroreport 16:1923-1927.

Stein BE, London R, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci 8:497-506.

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge, MA: MIT.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098-1114.

Teder-Salejarvi WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. J Cogn Neurosci 17:1396-1409.

Teder-Salejarvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106-114.

Vroomen J, de Gelder B (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. J Exp Psychol Hum Percept Perform 26:1583-1590.

Watkins S, Shams L, Josephs O, Rees G (2007) Activity in human V1 follows multisensory perception. Neuroimage 37:572-578.

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. Neuroimage 31:1247-1256.

Chapter 4: Effect of Attention on Early Cortical Processes underlying the Sound-induced Extra Flash Illusion

ABSTRACT

When a single flash of light is presented interposed between two brief auditory stimuli separated by 60-100 ms, subjects typically report perceiving two flashes (Shams et al., 2000, 2002). Using event related potential recordings (ERP), we previously found that perception of the illusory extra flash was based on a very rapid dynamic interplay between auditory and visual cortical areas that was triggered by the second sound (Mishra et al., 2007). In the current study we investigated the effect of attention, as well as field of stimulus presentation (upper vs. lower) on the ERP correlates of the sound induced illusory flash in fifteen individuals who perceived the cross-modal illusion. All early ERP components in the crossmodal difference wave associated with the extra flash illusion were significantly enhanced by selective spatial attention. The earliest attentional modulation was an amplitude increase of the positive-going PD110/ PD120 component, which has been previously shown to have a significant correlation with the frequency of illusory percepts amongst individuals (Mishra et al., 2007). The polarity of the early PD110/ PD120 component did not differ as a function of the field of stimulus presentation. This, along with the source localization of the component, suggested that its principal generator lies in extrastriate visual cortex. These results indicate that the neural processes underlying the extra flash illusion are attention dependent and not the result of an automatic cross-modal integration process.

INTRODUCTION

The natural world is multi-modal in its properties, and inter-sensory interactions in the brain are critical to the generation of coherent percepts and the control of subsequent behavior (reviewed in Stein and Meredith, 1993, Calvert, 2001, Amedi et al., 2005, Macaluso and Driver, 2005). Within the audio-visual domain, numerous behavioral studies have shown that simultaneous auditory and visual inputs interact such that visual perception can be altered by audition and vice versa. For example, the perceived location of sounds is robustly altered by concurrent visual stimuli at a nearby location, known as the ventriloquism effect (Pick et al., 1969, Bertelson, 1999, Hairston et al., 2003, Vroomen and de Gelder, 2004, Bonath et al., 2007). Conversely, simultaneous presentation of sounds can strikingly alter visual perception (Stein et al., 1996, Sekuler et al., 1997, Fendrich and Corballis, 2001, Recanzone, 2003, McDonald et al., 2003, 2005). One of the better studied visual illusions induced by audition is that introduced by Shams and colleagues (2000, 2002), wherein a single brief flash presented interposed between two pulsed sounds separated by 60-100 ms generates the percept of two distinct flashes, of which the second is illusory.

The neural basis of the extra illusory flash has been investigated in several physiological studies (Shams et al., 2001, 2005, Arden et al., 2003, Watkins et al., 2006, Mishra et al., 2007). Using ERP recordings Mishra et al. (2007) found that the cross-modal interactions underlying the illusory flash phenomenon had a complex but distinct neural signature. An important component was identified in the cross-modal interaction waveforms, designated the PD120 component as it was a positive deflection that peaked at 120 ms after the onset of the first sound. PD120 was found to originate within visual

cortex, and its amplitude in an individual was predictive of the frequency with which that person perceived the extra flash illusion. Indeed, the PD120 was completely absent in subjects who did not perceive the illusion. The objective of the present study was to investigate two factors that may modulate this early visual ERP component and other neural processes that underlie the sound-induced extra-flash illusion. These factors were selective spatial attention and location of the stimulus within the visual field.

Attention is known to amplify and enhance the processing of external stimuli to which it is allocated. Perceptual thresholds are lowered, reaction times are speeded and detection accuracy increases as a consequence of attention (e.g., Posner and Peterson, 1990, Luck et al., 1994, Carrasco, 2006). These behavioral effects have been linked with increases in the neural response to attended stimuli relative to unattended stimuli, ranging from increased firing rates at the single neuron level (Reynolds and Chelazzi, 2004) and enhancement of sensory ERPs (Hillyard and Anllo-Vento, 1998, Hopfinger et al., 2004) to increased blood flow within sensory cortices (Hopfinger et al., 2000, Kastner and Ungerleider, 2000, Kastner and Pinsk, 2004). Investigation of the role of attention in the context of cross-modal stimuli has only recently been initiated (Teder-Salejarvi et al., 1999, McDonald et al., 2001, 2003, Busse et al., 2005, Talsma and Woldorff, 2005, Molholm et al., 2007, Talsma et al., 2007a, b). These studies have found that attention may enhance cross-modal interactions as early as 100 ms after stimulus onset (Talsma and Woldorff, 2005), whereas the influence of attention spreads from one modality to another at later stages of processing beyond 200 ms (Busse et al., 2005, Molholm et al., 2007).

The role of attention in modulating the neural processing that underlies audio-visual illusions has been little investigated. Busse et al. (2005) studied the effect of attention in an experimental paradigm that simulated the ventriloquist illusion, but illusory perception was not measured as part of the study. The primary aim of the present study was to characterize the effect of attention on the audio-visual extra flash illusion discovered by Shams and colleagues (2000). To do so, stimuli were presented at two locations, one in the upper and one in the lower visual field, while subjects focused attention on only one of the locations at a time. The upper and lower visual field locations were chosen in order to help delineate the possible role of primary visual (striate) cortex in the generation of early cross-modal interaction components. It is well known that the earliest component of the visual evoked potential, the so-called C1 elicited at 50-90 ms, inverts in polarity for stimuli presented in the upper vs. the lower field, which supports the general consensus that the C1 originates in large part from striate cortex (Jeffreys, 1968, Clark et al., 1995, Martinez et al., 2001, Di Russo et al., 2002, 2003). Following this logic, any ERP components associated with the extra flash illusion that were primarily located within the striate cortex may also exhibit such polarity inversion.

MATERIALS & METHODS

Task and Stimuli

Fifteen right-handed healthy adults (8 females, mean age 21.4 yrs) participated in the study after giving written informed consent as approved by the University of California, San Diego Human Research Protections Program. Each participant had normal or corrected-to-normal vision and normal hearing. All subjects chosen for the

experiment perceived the sound-induced extra flash illusion as tested in a short 5 minute screen prior to the main experiment. Illusory perception threshold was set at 50% of the presented trials within the screen.

The experiment was conducted in a sound-attenuated chamber having a background sound level of 32 dB and a background luminance of 2 cd/m^2. Subjects maintained fixation on a central cross positioned at a viewing distance of 120 cm. Auditory (A) and visual (V) stimuli were delivered from paired speakers and red light emitting diodes (LED), one pair in the upper visual field (UVF) and another in the lower visual field (LVF). The speaker/LED pairs were positioned at 20º eccentricity to the left of fixation and at 30º polar angle above and below the horizontal meridian (Fig. 1A). The eccentricity of the stimuli was the same as that used previously (Mishra et al., 2007). Each visual stimulus was a 5 ms 75 cd/m^2 flash, and each auditory stimulus was a 10 ms 76 dB noise burst.

Six different stimulus combinations were presented one at a time to either the UVF or LVF in random order (Fig. 1B). Both the order of the combinations and field of presentation were randomized on each block of trials. The combinations included unimodal auditory stimuli, occurring in pairs (A_1A_2) and unimodal visual stimuli occurring singly (V_1) or in pairs (V_1V_2). Bimodal stimulus combinations included $A_1V_1A_2$ and $A_1A_2V_1$. In this terminology, suffixes 1 or 2 denote the first or second occurrence of the auditory or visual component of each stimulus combination. Finally, blank or no-stimulus (no-stim) trials ERPs were recorded over the same epochs as for actual stimuli but with no stimulus presented. The timing of the A and V components for each stimulus combination is shown in Fig. 1B. The SOA between the two stimuli in the

A_1A_2 and V_1V_2 pairs was 70 ms in every stimulus combination that included them. The SOA between A_1 and V_1 was 10 ms for $A_1V_1A_2$, and V_1 followed A_1 by 200 ms for $A_1A_2V_1$. The $A_1A_2V_1$ stimulus with the delayed flash did not produce an illusory second flash and thus served as a stimulus-matched behavioral control for the $A_1V_1A_2$ test stimulus that did produce the illusion, thereby ensuring that reports of the visual illusion were not based on simply counting the number of sounds.

Stimuli were presented in 16 blocks with 24 trials of each of the six stimulus combinations delivered in a randomized sequence (12 to the UVF and 12 to the LVF) on each block. All configurations occurred with equal probability and were presented at irregular intervals of 800-1200 ms. Within each block a 5 s break period of no stimulation was given every 30 s. Subjects were instructed to attend to the visual stimuli in either UVF or LVF on each block and report the number of flashes perceived (one or two) after each stimulus combination occurring in the attended field that contained one or two flashes. Subjects were instructed to ignore all stimuli in the unattended visual field, and no responses were required to the unimodal auditory stimuli. The order of attended blocks was counter-balanced across subjects. Overall, 192 trials were recorded for each attended as well as unattended stimulus combination in each visual field.

Electrophysiological (ERP) Recordings

The EEG was recorded from 62 electrode sites using a modified 10-10 system montage (Teder-Sälejärvi et al., 2005). Horizontal and vertical electro-oculograms (EOGs) were recorded by means of electrodes at the left and right external canthi and an

electrode below the left eye, respectively. All electrodes were referenced to the right mastoid electrode. Electrode impedances were kept below 5 kΩ.

All signals were amplified with a gain of 10,000 and a bandpass of 0.1-80 Hz (-12 dB/octave; 3dB attenuation) and were digitized at 250 Hz. Automated artifact rejection was performed prior to averaging to discard trials with eye movements, blinks or amplifier blocking. Signals were averaged in 500 ms epochs with a 100 ms pre-stimulus baseline. The averages were digitally low-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 46 Hz) to remove high frequency noise produced by muscle activity and external electrical sources. The filtered averages were digitally re-referenced to the average of the left and right mastoids.

The three-dimensional coordinates of each electrode and of three fiducial landmarks (the left and right pre-auricular points and the nasion) were determined by means of a Polhemus spatial digitizer (*Polhemus Corp., Colchester, VT*). The mean cartesian coordinates for each site were averaged across all subjects and used for topographic mapping and source localization procedures.

Cross-modal interaction difference waves were calculated for the $A_1V_1A_2$ stimulus that generated the percept of the illusory extra flash by subtracting the ERPs elicited by the individual unimodal components of the bimodal configuration from the ERP elicited by the total configuration. This difference wave was termed Ill_Diff as it reflected the neural activity associated with the illusory second flash; the Ill_Diff was separately calculated for stimuli in the upper and lower visual fields and for both the attended and unattended conditions, as follows:

$$\text{Ill\_Diff} = [(A_1V_1A_2) + \text{no-stim}] - [A_1A_2 + V_1]$$

Four such difference waves were calculated: for attended stimuli in the UVF ($Ill\_Diff_{ATT\text{-}UVF}$), for unattended stimuli in the UVF ($Ill\_Diff_{UNATT\text{-}UVF}$), for attended stimuli in the LVF ($Ill\_Diff_{ATT\text{-}LVF}$), and for unattended stimuli in the LVF ($Ill\_Diff_{UNATT\text{-}LVF}$).

The blank or no-stimulus (no-stim) trials were included in the calculation of these cross-modal difference waves to balance any prestimulus activity (such as a negative going anticipatory CNV) that may extend into the post-stimulus period and was present on all trials. If the no-stim trials were not included such activity would be added once but subtracted twice in the difference wave, possibly introducing an early deflection that could be mistaken for a true cross-modal interaction (Teder-Sälejärvi et al., 2002; Talsma & Woldorff 2005; Gondan & Röder 2006).

The attention effects in each visual field were calculated by subtracting the unattended Ill_Diff wave from the attended Ill_Diff wave in UVF as well as LVF to yield attention double difference waves, as follows:

Ill_Diff_{UVF} double difference = $Ill\_Diff_{ATT\text{-}UVF} - Ill\_Diff_{UNATT\text{-}UVF}$

Ill_Diff_{LVF} double difference = $Ill\_Diff_{ATT\text{-}LVF} - Ill\_Diff_{UNATT\text{-}LVF}$

Data Analysis

ERP components observed in each Ill_Diff difference wave were first tested for significance with respect to the prestimulus baseline and compared by t-tests over all subjects (n=15). For all analyses difference wave components were quantified as mean amplitudes within specific latency windows around the peak for each identified positive difference (PD) or negative difference (ND) component with respect to the mean voltage

of a 100 ms prestimulus baseline. Components in the Ill_Diff$_{UVF}$ difference wave (both attended and unattended) were measured at 112-132 ms (PD120), 164-184 ms (PD180), and 240-260 ms (ND250). For the Ill_Diff$_{LVF}$ difference wave (again, both attended and unattended) components were measured at 104-124 ms (PD110), 164-184 ms (PD180), and 228-248 ms (ND240). Each of these components was measured as the mean voltage over a specific cluster of electrodes where its amplitude was maximal. The PD120 and PD110 components for stimuli in UVF and LVF respectively, were measured over 15 occipital electrode sites (6 in each hemisphere and 3 over midline), PD180 amplitudes (for both UVF and LVF) were measured over fronto-central electrode clusters (8 in each hemisphere and 4 over midline), and the ND250/ND240 components were measured over a similar set of central electrodes (8 in each hemisphere). Components in the attention double difference waves were also characterized over the same electrode clusters as their counterparts in the Ill_Diff difference waves.

Scalp distributions of ERP components in the Ill_Diff difference waves were compared after normalizing their amplitudes prior to ANOVA according to the method described by McCarthy and Wood (1985). For posteriorly distributed components (PD120/PD110) comparisons were made over 18 occipital electrode sites (7 in each hemisphere and 4 over midline). For the other components (PD180 and ND250/ND240) comparisons were made over 38 electrodes spanning frontal, central, parietal and occipital sites (15 in each hemisphere and 8 over midline). Differences in scalp distribution were reflected in significant stimulus condition (ATT vs. UNATT or UVF vs. LVF) by electrode interactions.

Modeling of ERP Sources

Source localization was carried out to estimate the intracranial generators of components in the grand-averaged difference waves within the same time intervals as those used for statistical testing. Source locations were estimated by dipole modeling using BESA (Brain Electrical Source Analysis 2000, version 5). The BESA algorithm estimates the location and the orientation of multiple equivalent dipolar sources by calculating the scalp distribution that would be obtained for a given dipole model (forward solution) and comparing it to the actual scalp-recorded ERP distribution (Scherg, 1990). The algorithm interactively adjusts (fits) the location and orientation of the dipole sources in order to minimize the relative variance (RV) between the model and the observed spatio-temporal ERP distribution. This analysis used the three-dimensional coordinates of each electrode site as recorded by a spatial digitizer. Symmetrical pairs of dipoles were fit sequentially to the components of interest; dipole pairs were constrained to be mirror-symmetrical with respect to location but were free to vary in orientation.

To visualize the anatomical brain regions giving rise to the different components the locations of BESA source dipoles were transformed into the standardized coordinate system of Talairach and Tournoux (1988) and projected onto a structural brain image supplied by MRIcro (Rorden and Brett, 2000) using AFNI (Analysis of Functional NeuroImaging: Cox, 1996) software.

Trial based Analysis

A trial-by-trial analysis of the ERPs elicited in association with the illusory second flash in the UVF and LVF attended Ill_Diff waveforms was performed by

separating the $A_1V_1A_2$ trials on which subjects reported seeing the illusory second flash, ($Ill\_Diff_{ATT\text{-}UVF/SEE}$ and $Ill\_Diff_{ATT\text{-}LVF/SEE}$ trials) from trials on which the illusion was not perceived ($Ill\_Diff_{ATT\text{-}UVF/NOSEE}$ and $Ill\_Diff_{ATT\text{-}LVF/NOSEE}$ trials). To isolate ERP activity associated with seeing the illusory second flash, the double difference waves, $Ill\_Diff_{ATT\text{-}UVF/SEE\text{-}NOSEE}$ and $Ill\_Diff_{ATT\text{-}LVF/SEE\text{-}NOSEE}$ were generated for every subject. The main component in these two double difference waves was measured at 136-160 ms (ND150). This components was quantified as the mean voltage over the same fronto-central electrode clusters as those used to measure PD180 in the Ill_Diff waveforms (see Data Analysis section).

Frequency domain Analysis

To analyze differences in oscillatory cortical activity between trials on which the illusion was seen vs. not seen, the single trial EEG signal on each channel for $A_1V_1A_{2ATT\text{-}UVF/SEE}$, $A_1V_1A_{2ATT\text{-}UVF/NOSEE}$, $A_1V_1A_{2ATT\text{-}LVF/SEE}$ and $A_1V_1A_{2ATT\text{-}LVF/NOSEE}$ trials was convolved with Morlet wavelets in a 2 sec window centered at stimulus onset. Instantaneous power and phase were extracted at each time point over 91 frequency scales from 0.6 to 101 Hz incremented logarithmically (Lakatos et al., 2005). The square root of the power values were averaged over single trials to yield the total average spectral amplitude (in uV). The average spectral amplitude at each time point and frequency was baseline corrected by subtracting the average spectral amplitude in the -300 to -50 ms pre-stimulus interval (corrected separately for each frequency band) (Tallon-Baudry et al., 1998). The phase locking index across trials was calculated by

normalizing the complex wavelet decomposition on every trial by its absolute value, and averaging this quantity over all trials.

Significant differences in spectral amplitude between trials on which the illusion was seen vs. not seen were tested within the 20-50 Hz frequency range over the occipital scalp in order to investigate spectral modulations similar to those previously reported (Mishra et al., 2007). Specific differences were analyzed using ANOVA (Kiebel et al., 2005) within an interval spanning the illusion seen vs. not-seen differences at 90-120 ms in the 25-35 Hz frequency range.

RESULTS

Behavioral results

Subjects indicated by pressing one of two buttons the number of flashes perceived (one or two) in each stimulus combination in the attended field that contained flashes. Mean percentages of responses on which two flashes were reported over all 15 subjects are given in Fig. 2 and corresponding reaction times in Table 1. Subjects reported perceiving an illusory second flash on an average of 47% and 45% of the $A_1V_1A_2$ attended trials in UVF and LVF respectively. Subjects responded accurately to both unimodal visual stimuli (V_1 and V_1V_2) and to the bimodal control stimulus ($A_1A_2V_1$). For all stimuli, there were no significant differences in behavioral performance between stimuli presented in the upper vs. the lower visual field either for detection rates (UVF vs. LVF : $F(1,14) = 1.79$, p = n.s.) or for reaction times (UVF vs. LVF : $F(1,14) = 2.17$, p = n.s.).

Reaction times differed significantly across stimulus conditions ($F(3,42) = 17.98$, $p < 0.0001$). Reaction times on unimodal double flash trials were found to be significantly faster than reaction times on single flash trials in both visual fields (V_1V_2 vs. V_1: $F(1,14) = 27.30$, $p < 0.0002$). Reaction times on bimodal $A_1V_1A_2$ trials on which two flashes were perceived vs. when only a single flash was seen were not significantly different overall ($F(1,14) = 0.55$, p = n.s.). For $A_1V_1A_2$ stimuli in the lower field, however, a trend similar to unimodal flash trials with faster reaction times on illusory trials (on which two flashes were seen) was observed (Visual Field x Illusory trials interaction: $F(1,14) = 4.96$, $p < 0.05$)

ERP Results

The grand-averaged ERPs (over all 15 subjects) elicited by the attended as well as unattended illusion-inducing $A_1V_1A_2$ stimulus and by its unimodal components, V_1 and A_1A_2 are shown for presentations in the UVF and LVF in Fig. 3 and 4, respectively. The unimodal stimuli showed the typical pattern of ERP components. Visual ERPs to V_1 had characteristic P1 (120 ms), N1 (180 ms) and P2 (200 ms) components with maxima at posterior electrode sites, and an earlier N1 (165 ms) at anterior sites. The ERPs to A_1A_2 included auditory-evoked P1 (60 ms), N1 (105 ms) and P2 (180 ms) components with maxima at fronto-central electrode sites. It should be noted that the sharp positive-going deflection that peaks at around 20 ms was produced by the sound evoked post-auricular (P.A.) muscle reflex (Picton et al., 1974) recorded at the mastoid reference site.

For the unimodal stimuli, attention effects on ERP components were only found within the visual modality, with both the P1 and N1 components being enlarged to visual

stimuli in the attended field (see supplementary table S1). Scalp distributions and source localizations of the attention effects on visual P1 and N1 components are characterized in the supplementary section (see supplementary Fig. S1) and were consistent with visual attention effects found in previous studies (Martinez et al., 2001, 2006, 2007, Di Russo et al., 2002, 2003). It was not surprising that no attention effects were found on auditory ERPs, as the task demands did not require paying attention to the auditory modality. For the bimodal $A_1V_1A_2$ stimuli, the attended ERPs showed a larger positivity relative to the unattended waveforms within the 120-150 ms time range over occipital electrode sites. The attention differences for the bimodal stimuli were not further characterized in these ERPs, however, as the effect of attention on the auditory and visual components of the configuration could not be separated from the attention effects on the cross-modal interaction of the unimodal components. Hence, in subsequent analyses cross-modal difference waves were calculated (see Methods), and attention effects on the cross-modal interaction components therein were analyzed.

The cross-modal interaction difference waves for the $A_1V_1A_2$ stimuli were calculated for both upper and lower visual fields and for both attended and unattended conditions (Fig. 5). For the attended difference waves, $Ill\_Diff_{ATT\text{-}UVF}$ and $Ill\_Diff_{ATT\text{-}LVF}$, the earliest significant components were prominent positivities at occipital sites that extended over the interval 100-150 ms. These positivities were quantified around their early peaks, PD120 in the 112-132 ms time interval for UVF and PD110 in the 104-124 ms latency for LVF, respectively. The PD120/PD110 deflections were followed by a larger positivity peaking at 180 ms over anterior sites in both attended UVF and LVF waveforms, termed PD180. The PD180 also extended posteriorly to the O1/O2 sites at

reduced amplitude. The final components characterized within the attended Ill_Diff waves were negativities within the 240-260 ms interval (ND250) in UVF and the 228-248 ms interval (ND240) in LVF. The amplitudes and significance of these components with respect to the pre-stimulus baseline are given in Table 2. For the PD120/ PD110 components a significant correlation with behavior was not found (UVF: $r(13) = 0.11$, p = n.s., LVF: $r(13) = 0.21$, p = n.s.). As in our previous study, components occurring after 300 ms were not analyzed because of the likelihood that neural activity related to decision making and response preparation would be confounded with activity related to cross-modal interaction and perceptual processing (Mishra et al., 2007).

The difference wave components in the unattended waves, $Ill\_Diff_{UNATT\text{-}UVF}$ and $Ill\_Diff_{UNATT\text{-}LVF}$, were characterized in the same time intervals as the components in the attended difference waveforms. The amplitudes and significance levels of these components relative to baseline are also given in Table 2. The early PD120/PD110 components did not reach significance in the unattended waveforms, while the later PD180 and ND250/ND240 components were significantly diminished relative to their attended counterparts. PD180 in the $Ill\_Diff_{UNATT\text{-}LVF}$ waveforms and ND250 in the $Ill\_Diff_{UNATT\text{-}UVF}$ waves also did not reach statistical significance.

The scalp voltage distributions of the attended and unattended Ill_Diff wave components in UVF and LVF are shown in Fig. 6. Both PD120 and PD110 attended components in UVF and LVF respectively, had occipital scalp distributions. The right hemispheric preponderance of PD120 in the $Ill\_Diff_{ATT\text{-}UVF}$ wave did not reach significance ($F(1,14)=3.24$, p = n.s.), nor did the slight left laterality of PD110 in $Ill\_Diff_{ATT\text{-}LVF}$ ($F(1,14)=3.36$, p = n.s.). The topographies of the unattended components

in the PD120/PD110 latency ranges are shown (Fig. 6A), but their amplitudes did not reach statistical significance, as mentioned above. For the PD120/PD110 components there was a significant effect of attention for both UVF and LVF stimuli (graphed in Fig. 6A). The subsequent PD180 had a fronto-central distribution in the $Ill\_Diff_{ATT\text{-}UVF}$, $Ill\_Diff_{ATT\text{-}LVF}$ and $Ill\_Diff_{UNATT\text{-}LVF}$ difference waves with a non-significant right hemispheric preponderance. For the $Ill\_Diff_{UNATT\text{-}UVF}$ waveform, however, the topography of PD180 was shifted posteriorly to centro-parietal sites. Although we have no explanation for this shift in scalp distribution, we note that a similar shift in topography was also reported in a previous study comparing attended and unattended cross-modal difference waves (Talsma and Woldorff, 2005). The attention effect on PD180 was significant for both UVF and LVF stimuli (Fig. 6B). Lastly, the ND250 (UVF) and ND240 (LVF) components had prominent fronto-central distributions with a significant attention effect in both visual fields (Fig. 6C).

The scalp topographies of the above Ill_Diff wave components were statistically compared between the attended and unattended conditions and for the two visual field locations following normalization according to the method of McCarthy & Wood (1985). Comparisons were made in an omnibus ANOVA with visual field location (UVF vs. LVF), attention (ATT vs. UNATT) and electrodes as factors for each difference wave component. The topographies of the attended PD120 and PD110 components in UVF and LVF respectively, were found to differ from one another (Visual Field x Electrode interaction: $F(17, 238) = 2.73$, $p < 0.0004$). The attended PD120/ PD110 topographies differed from their unattended counterparts (Attention x Electrode interaction: $F(17, 238) = 3.57$, $p < 0.0001$), which was expected as the scalp distributions of the latter were at

noise levels. However, no differences in spatial topography were found with respect to attention exclusive to one or the other visual field as reflected by the non-significant three way interaction of Visual Field x Attention x Electrodes ($F(17,238) = 0.51$, p = n.s.). For the PD180 component, the spatial topography of the component in the $Ill\_Diff_{UNATT\text{-}UVF}$ wave differed from the other three PD180 topographies (Fig. 6B) as reflected in the significant Visual Field x Attention x Electrode interaction ($F(37, 518) = 1.66$, $p < 0.02$) and the significant post-hoc Attention x Electrode interaction ($F(37, 518) = 2.01$, $p < 0.0006$) in the upper visual field. Other post-hoc topography comparisons for the PD180 with respect to attention in the lower visual field, and comparison of attended topographies in the two visual fields did not differ. Lastly, the ND250 component did not show topographic differences either with respect to visual field (Visual Field x Electrode interaction: $F(37, 518) = 1.27$, p = n.s.) or attention (Attention x Electrode interaction: $F(37, 518) = 1.35$, p = n.s.).

In order to characterize further the effects of attention on these cross-modal interactions, double difference waves were calculated for each visual field subtracting the unattended from the attended Ill_Diff waveforms (see Methods). Fig. 7 shows the waveforms resulting from this subtraction. All components characterized in the Ill_Diff difference waves above could be identified as significant components in the double difference waves: $PD120_{diff}$, $PD180_{diff}$ and $ND250_{diff}$ in the Ill_Diff_{UVF} double difference wave and $PD110_{diff}$, $PD180_{diff}$ and $ND240_{diff}$ in the Ill_Diff_{LVF} double difference wave. The amplitudes of these components and their significance with respect to the pre-stimulus baseline (i.e., the significance of the attention effect) are provided in Table 3.

The size of the attention effects in the upper vs. lower visual fields did not differ for any component.

The scalp topographies of the double difference wave components are shown in Fig. 8. The distributions of the attention effects on each component were compared in the UVF vs. LVF following normalization according to the method of McCarthy & Wood (1985). The voltage topography of $PD120_{diff}$ (UVF) was found not to significantly differ from that of $PD110_{diff}$ (LVF) (Visual Field x Electrode interaction: $F(17, 238) = 0.67$, p = n.s). However, the later components did differ in topography in the UVF vs. LVF ($PD180_{diff}$: Visual Field x Electrode interaction: $F(37, 518) = 1.84$, $p = 0.003$, and $ND250/240_{diff}$: Visual Field x Electrode interaction: $F(37, 518) = 1.59$, $p = 0.02$).

Source Analysis

The neural generators of the significant components identified in the Ill_Diff attended and unattended difference waves as well as the attention double difference waves in the upper and lower visual field locations were modeled using dipole source localization. Pairs of dipoles were fit to the scalp topographies of the components using the BESA algorithm (Scherg, 1990). The location of the BESA dipoles were transformed into the standardized coordinate system of Talairach and Tournoux (1988) and superimposed on the rendered cortical surface of a single individual's brain (Fig. 9). Talairach coordinates of the dipole pairs and an estimate of their goodness of fit as reflected by residual variance are listed in Table 4 and 5.

In the upper visual field, the earliest components, PD120 in the $Ill\_Diff_{ATT\text{-}UVF}$ difference wave and $PD120_{diff}$ in the Ill_Diff_{UVF} double difference wave, were both

localized to ventral-lateral extrastriate cortex in the region of the fusiform gyrus (Fig. 9A). The dipole in the right hemisphere accounted for greater component variance than the left hemisphere dipole. In the lower visual field, PD110 and PD110diff, the corresponding attended and attentional difference components in the $Ill\_Diff_{ATT-LVF}$ difference wave and the Ill_Diff_{LVF} double difference wave, respectively, also localized to lateral extrastriate visual cortex approximately 10-15 mm dorsal to the $PD120/PD120_{diff}$ dipoles. The left and right hemisphere dipoles accounted for equivalent variance in the case of the LVF dipole fits. The later components, both PD180 and ND250/ND240 in the attended, unattended and attentional difference waveforms were consistently localized to the region of the superior temporal polymodal cortex (Fig. 9B, C). There were no marked differences in the dipole localizations for these later components between stimuli in the upper versus lower visual field locations. The variability in scalp distributions for the same component in the attended, unattended and attentional difference waveforms (Fig. 6 and 8) could be accounted for by different orientations of the underlying source dipoles, as opposed to differences in dipole location.

Trial based Analysis

A trial-by-trial analysis was conducted that compared the $Ill\_Diff_{ATT-UVF}$ and the $Ill\_Diff_{ATT-LVF}$ waveforms on trials where the extra flash illusion was reported (SEE trials) versus those on which it was not seen (NO-SEE trials). The SEE and NOSEE difference waves in both UVF (Fig. 10A) and LVF (Fig. 10B) differed significantly from each other in the 136-160 ms latency, as evident in an overall ANOVA over both VFs ($F(1,14) = 9.19$, $p < 0.009$). These differences could be seen in the SEE-NOSEE trials

difference waveforms as a negative component peaking at 150 ms (ND150), which was significant in UVF ($t(1,14) = 2.31$, $p< 0.04$) as well as in LVF ($t(1,14) = 2.67$, $p< 0.02$). The ND150 components in both UVF and LVF had amplitude maxima over fronto-central sites, while no differences between trials were found to be significant over occipital electrodes. The scalp topographies of the ND150 components in UVF and LVF are shown in Fig. 11A. The generators of the ND150 in the UVF were localized using BESA dipole fits to the anterior region of the superior temporal gyrus, while the LVF dipoles were relatively more posterior in the superior temporal area in the vicinity of the auditory cortex (Table 5, Fig. 11 B).

Frequency domain Analysis

Differences in oscillatory cortical activity between attended SEE and NO-SEE trials were analyzed over all subjects in both visual field locations using a Morlet wavelet decomposition in the time-frequency domain. A very early burst of enhanced spectral amplitude (or enhanced power: EP) was observed at 25-50 Hz on all channels; this effect could be attributed to the short latency (10-15 ms) reflex contraction of the post-auricular muscle affecting the mastoid reference electrode ("post-auricular reflex" in Fig. 12). SEE vs. NO-SEE trial differences were specifically compared at occipital electrodes where robust differences had been previously observed in the 25-35 Hz frequency range at a latency of 100-150 ms (Mishra et al., 2007). Over this frequency range and the 90-130 ms time interval, spectral amplitude differences could be qualitatively seen between SEE and NO-SEE trials over both UVF and LVF, with greater cumulative spectral power over SEE trials. The spectral amplitude within the region of interest was significant for SEE

trials at the occipital O2 electrode shown in Fig. 12 for both UVF (Amp. = 0.42 ± 0.18, $t(14) = 2.28$, $p < 0.04$) and LVF (Amp. = 0.45 ± 0.17, $t(14) = 2.35$, $p < 0.04$) but not for NO-SEE trials (UVF: Amp. = 0.32 ± 0.17, $t(14) = 1.81$, p = n.s.; LVF: Amp. = 0.25 ± 0.17, $t(14) = 1.46$, p = n.s.). However, the difference between the SEE and NO-SEE spectral amplitudes over an occipital electrode cluster as defined in our previous study (Mishra et al., 2007) was not significant (UVF: $F(1,14) = 0.002$, p = n.s., LVF: $F(1,14) = 0.02$, p = n.s). This insignificant result may be attributed to the low signal to noise levels in the SEE/ NO-SEE trial sets consisting of fewer than 100 trials per set per subject in the present experiment.

DISCUSSION

The present study investigated the effect of attention on neural interactions that underlie the cross-modal (audio-visual) extra flash illusion. Subjects perceived the cross-modal $A_1V_1A_2$ stimulus as containing an illusory second flash on an average 44-46% of attended trials in both the upper and lower peripheral visual field locations. The cross-modal interaction underlying the extra flash illusion was calculated by subtracting the ERPs elicited by the unimodal components (V_1 and A_1A_2) from the ERPs to the cross-modal combination ($A_1V_1A_2$) for each spatial location and attended state. This interaction difference wave associated with the illusion was termed Ill_Diff. For both upper and lower visual field stimuli, the components identified in Ill_Diff included PD120 (PD110 in the lower field) that originated from ventral occipito-temporal extrastriate visual cortex, followed by PD180 and ND250 (ND240 in lower field), both of which had sources within superior temporal polymodal cortex. These difference wave components

were the same as those previously characterized in a study of the neural basis of the illusion (Mishra et al., 2007). In the present study, PD120/ PD110 were not found to invert in polarity for upper vs. lower visual field stimulus presentations, confirming that the principal generators of these components lay outside of striate cortex. The PD120/ PD110 components were also found to be attention dependent, as they were reduced to non-significant levels in the non-attended visual field. The trial-by-trial analysis of the attended Ill_Diff waveforms revealed an enlarged negativity, ND150 with underlying sources within superior temporal cortex, that differentiated trials on which the extra flash illusion was seen vs. not seen. The illusion was thus found to depend on a sequence of neural events that included an early attention-dependent enhancement in extrastriate visual cortex followed by a trial specific modulation within superior temporal area that accompanied the illusory percept.

The presence of the early PD120 (PD110) component in the illusory difference waveforms replicated the observations in our previous study (Mishra et al., 2007) that characterized the neural correlates of the sound-induced extra flash illusion. In that study the amplitude of the early PD120 component was found to positively correlate with the proportion of trials on which an individual subject perceived the illusion. Since individuals who did not perceive the illusion were excluded from the present study, the PD120 amplitude was not found to fluctuate much across subjects and hence, a significant correlation with behavior could not be shown. The present study revealed a new aspect of the PD120, that attention is a prerequisite for the generation of this early visual component. When the $A_1V_1A_2$ stimulus was ignored in either visual field, the PD120 component was not elicited in the interaction difference waves. This is the first

report to our knowledge, of an attention effect on cross-modal interactions in sensory specific cortex. In our previous study (Mishra et al., 2007), although explicit instructions to attend were not provided, subjects needed to maintain attention to all stimuli in order to make perceptual judgments for every stimulus that contained a flash. Taken together, our previous and current investigations suggest that the PD120 component is prevalent amongst subjects who frequently perceive the extra-flash illusion and when attention is specifically directed towards the multi-modal stimuli.

The neural generators of the PD120 component in the upper as well as lower visual field (PD110) were found to localize to ventral lateral extrastriate visual cortex near the fusiform gyrus. The PD110 to lower field stimuli localized about 10-15 mm dorsal to the PD120 generators for upper field stimuli. This difference might be accounted for by differential activation of retinotopic visual areas as a function of stimulus location, but this cannot be confirmed because the dipole fitting technique used to estimate the sources has spatial localization error within this range. The PD120/ PD110 generators in the present study were also found to lie about 15 mm more anterior to the PD120 source foci in the previous study (Mishra et al., 2007). This difference might be attributed to the different source localization techniques used in the two studies (a distributed minimum-norm approach in Mishra et al. (2007) and dipole modeling in the present analysis) or simply to localization error. Alternatively, the new variables in the present study: selective spatial attention and stimulus location in the upper/ lower fields instead of at the horizontal meridian, may also have contributed to the localization differences between the two studies. In any case, the lack of a polarity inversion of the

PD120 component in the upper vs. lower field comparison in the present study provides strong evidence that its predominant generator site lies outside of striate cortex.

Since the PD120 (/PD110) is a rapidly generated component, emerging within 30-60 ms after onset of the second sound (A_2) within the $A_1V_1A_2$ stimulus, we previously hypothesized that direct connections between auditory and visual areas may be responsible for its generation (Mishra et al., 2007). Such connections have been characterized in recent years in anatomical labeling studies in primates (Falchier et al., 2002, Rockland and Ojima, 2003, Clavagnier et al., 2004) and have been shown to be denser in visual areas higher in the visual hierarchy than primary cortex. This anatomy is consistent with the localization of the PD120 to lateral extrastriate visual cortex. Furthermore, it is unlikely that the PD120 component could be driven via feedback from higher areas such as multisensory (superior temporal) cortex, given that feedback usually has a slower time course and there was no ERP evidence to suggest modulation of polymodal cortex prior to the PD120/ PD110. Interestingly, the attended PD120 (/PD110) within the present study localized to an extrastriate cortical region within the ventral visual stream where activity has been shown to be modulated by attention in numerous ERP and fMRI studies (Kastner et al., 1998, Corbetta et al., 2000, Hopfinger et al., 2000, Martinez et al., 2001, 2006, 2007). This suggests that the neural populations in visual cortex that are modulated by top-down attention are also the substrates of direct input from cross-sensory (auditory) cortices. In any case early cross-modal modulation highlights the intimate link between processing in different sensory streams as is being increasingly found in studies of multisensory integration (Schroeder and Foxe, 2005, Driver and Noesselt, 2008).

Beyond the PD120 component, two large components were found within the interaction difference waveforms, PD180 and ND250 (ND240 in the lower field), as had been characterized in our previous study (Mishra et al., 2007). The attended and unattended counterparts of these components as well as the attention effects (attended minus unattended component) were consistently best fit by dipole pairs in the superior temporal region. Some of these components, however, such as the PD180 for upper field stimuli showed distinctly different spatial topographies in the attended vs. unattended waveforms. The unattended PD180 in this case had a centro-parietal distribution different from the fronto-central distribution for the attended PD180. We have no current explanation for this shift in topography with attention, we note that a similar attention related shift was seen for the same component in a previous study of attention to audio-visual stimuli (Talsma and Woldorff, 2005). That the dipole fits to these differing scalp distributions were localized in close proximity to one another in superior temporal area can be accounted for by the very different orientations observed for the dipole sources. Alternatively, as 10-15% of the residual variance was not accounted for by the source models of the late components, it is possible that additional cross-modal brain regions, such as the inferior frontal region that was previously found as a secondary source for the PD180 (Mishra et al., 2007) or parietal cortex (reviewed in Driver and Noesselt, 2008), could be involved to some extent in generating the different topographies. Future studies using functional imaging in conjunction with ERPs could help delineate the underlying anatomical bases of these subtle differences in the effects of attention on multisensory interactions.

In our previous study (Mishra et al., 2007), the later components were not found to be correlated with illusory perception either between subjects or on a trial by trial basis. Further, the later components were present in the interaction difference waves to other cross-modal control stimuli in that study (A_1V_1 and $A_1V_1A_2V_2$), and have been previously noted in many other cross-modal ERP investigations (Teder-Sälejärvi et al., 2002, 2005; Molholm et al., 2002; Talsma & Woldorff 2005). These findings, along with the source localization of the later components to the superior temporal area which is a ubiquitous polymodal region (Calvert, 2001), corroborate the notion that PD180 and ND250 (/ND240) reflect very general aspects of cross-modal interaction.

In the present study the later components were also found to be manipulable by attention, with reduced amplitudes in the unattended waveforms. These results suggest that attention can significantly affect the process of multisensory integration and is in line with many previous investigations of the interaction between attention and cross-modal processing (Hillyard et al., 1984, Eimer and Schroger, 1998, Teder-Sälejärvi et al., 1999, Mc Donald et al., 2001, 2003, Talsma and Kok, 2002, Talsma and Woldorff, 2005). The present results most closely parallel the ERP findings by Talsma and Woldorff (2005), who studied the effects of attention on crossmodal interaction difference waves to simple audio-visual stimuli. Their experimental paradigm required attending to both auditory and visual components of an audio-visual stimulus at a given spatial location. In our task subjects did not need to pay attention to the auditory stimulus component, as only flashes were counted on the attended audio-visual trials. Yet the similarity of the electrophysiological results in these two studies suggests that visual spatial attention

(rather than auditory attention) was predominantly modulating the cross-modal integration processes.

Overall attention was found to influence both early and late stages of cross-modal processing associated with the audio-visual extra flash illusion, starting at around 100 ms after the onset of the visual stimulus. This is consistent with the sensory gain control mechanism of visual-spatial attention (Hillyard et al., 1998), which modulates the early P1 component (80-120 ms) as well as later components. The present results also imply that the extra flash illusion may not be an automatic multisensory integration process. This is in contrast with the ventriloquist effect, an auditory illusion influenced by vision that is reportedly not modulated by spatial attention (Bertelson, 1999, Bertelson et al., 2000, Vroomen et al., 2001a,b). For audio-visual stimulus configurations that induce the ventriloquist effect, however, ERPs to the auditory stimulus are enhanced when the associated visual stimulus is attended (Busse et al., 2005). Thus, at the neural level these reciprocal cross-modal illusions are both subject to modulation by attention.

In the trial by trial analysis, the cross-modal interaction Ill_Diff waves revealed an enlarged negativity (ND150) on trials when the illusion was perceived vs. not perceived. The ND150 was localized to the anterior region of the superior temporal cortex for stimuli in the upper field and somewhat more posteriorly (near the auditory cortex) for stimuli in the lower field. The trial by trial analysis did not yield any ERP differences over occipital cortex. In our previous study (Mishra et al., 2007), difference wave components were similarly found on trials where the illusion was perceived within a latency range of 100-150 ms (ND110 and ND130). These components also had neural generators localized to the superior temporal gyrus in the vicinity of auditory cortex, and

again no trial-specific ERP modulation of visual cortex was found. While the present trial-by-trial results generally replicate the previous findings, there are minor differences in timing and source localization that could be due to the many differences between the studies such as change in task demands from general attention to selective attention, different stimulus positions, and different source localization procedures. It should also be noted that the signal to noise ratio in the present study was nearly half of that in the previous study, as ERPs were recorded for both attended and unattended trials within the same experimental time constraints. Thus the data recorded in the present study was not particularly well-suited for a statistically robust trial by trial analysis. The lower signal to noise ratios of the present study may also have contributed to the failure to find frequency domain modulations of gamma power over occipital sites that were qualitatively observed but did not attain statistical significance as in our previous study (Mishra et al., 2007).

In summary, we found that the multisensory integration processes underlying the cross-modal auditory-induced extra flash illusion are significantly enhanced by selective spatial attention. The earliest modulation by attention was found at 100-130 ms within ventral occipito-temporal extrastriate visual cortex on a component (PD120/ PD110) that was previously shown to predict the frequency with which individuals perceived the illusion (Mishra et al., 2007). The generation of this critical component was found to be attention dependent in that no PD120 was generated on unattended trials. Concurrent with the latency range of the PD120, illusory trials were associated with a modulation of superior temporal activity (136-160 ms). These findings suggest that the illusion is the result of a rapid interplay between unisensory and multisensory cortical sites, and that

these neural processes are crucially attention dependent and not the result of an automatic integration process.

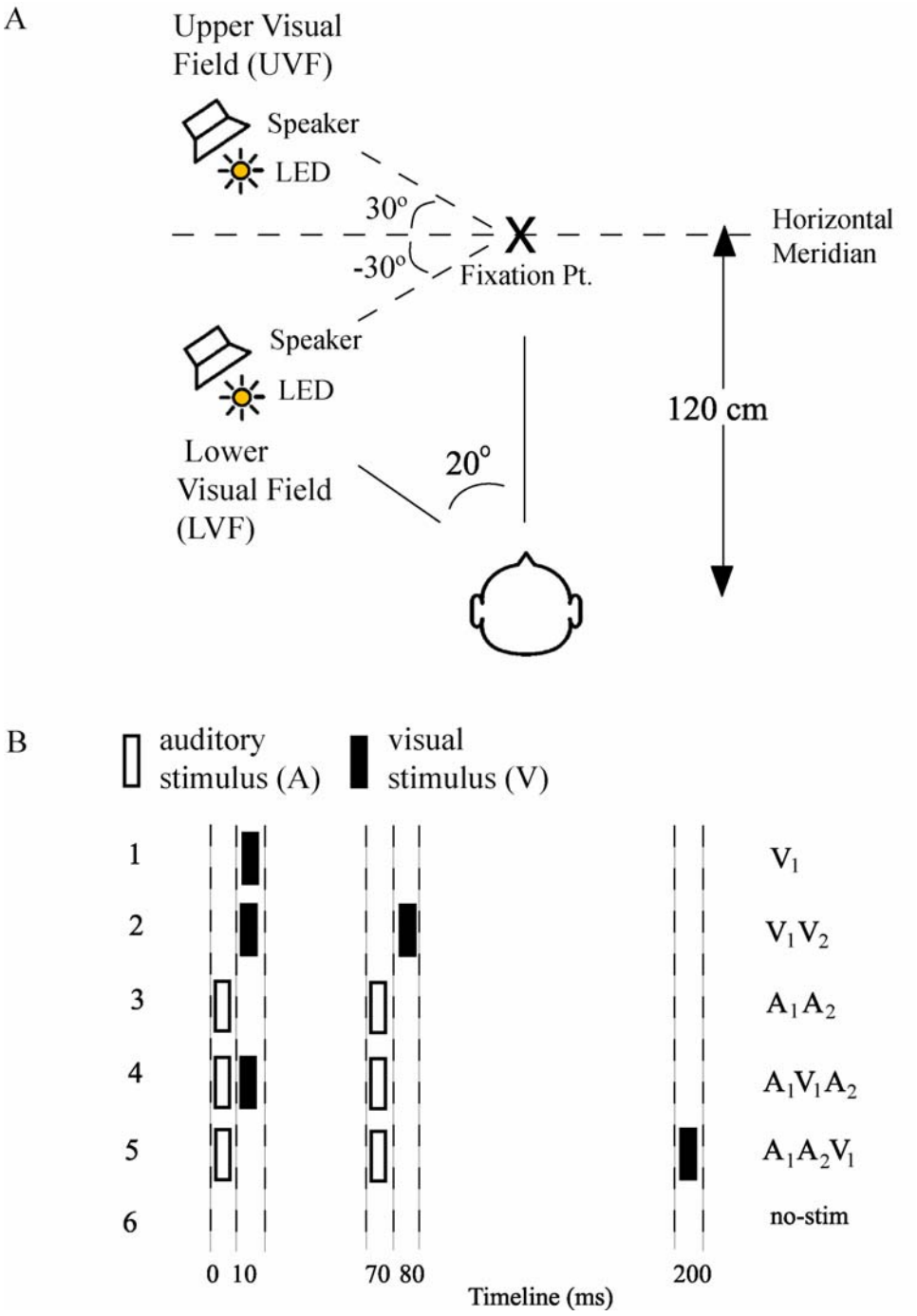

Figure 4.1 Overview of experimental design [A] Schematic diagram of experimental set-up [B] Listing of the six different stimulus configurations, which were presented to either the upper or lower visual field. Both the order of stimuli and the field of stimulation were randomized. Abscissa indicates times of occurrence of auditory (open bars) and visual (solid bars) stimuli. Auditory (A) and visual (V) stimuli are labeled 1 or 2 to designate their first or second occurrence in each configuration.

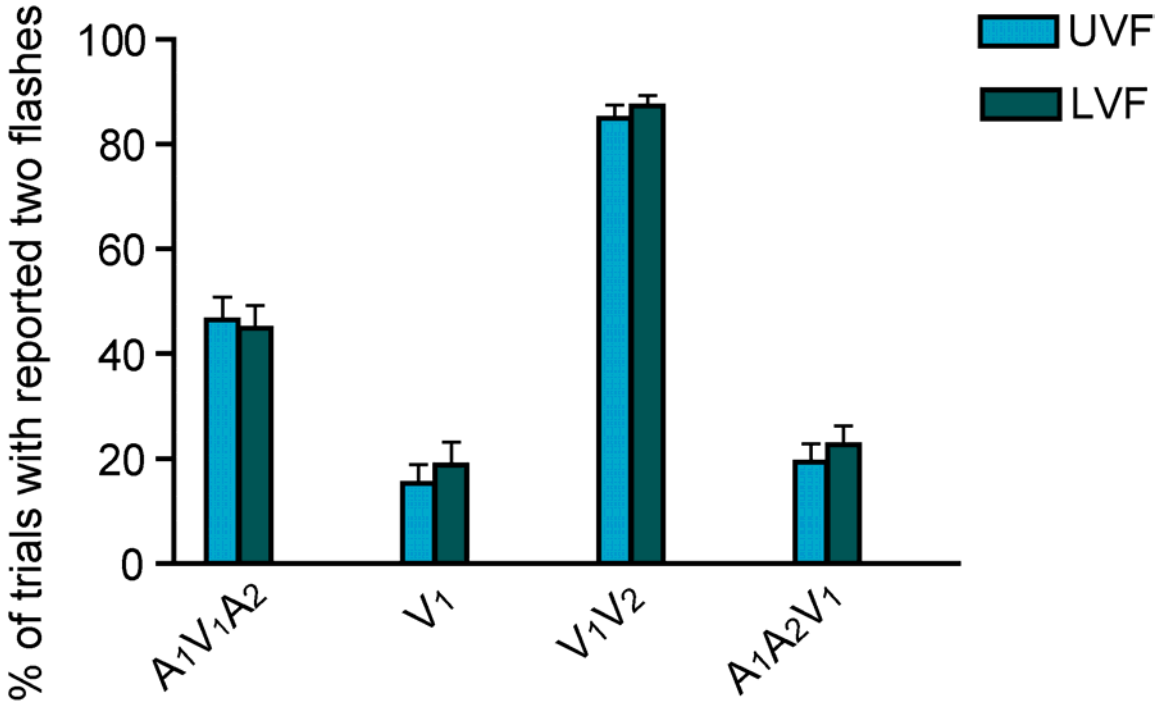

Figure 4.2 Behavioral performance comparisons in the upper (UVF) and lower visual fields (LVF) for all experimental stimuli that contained flashes.

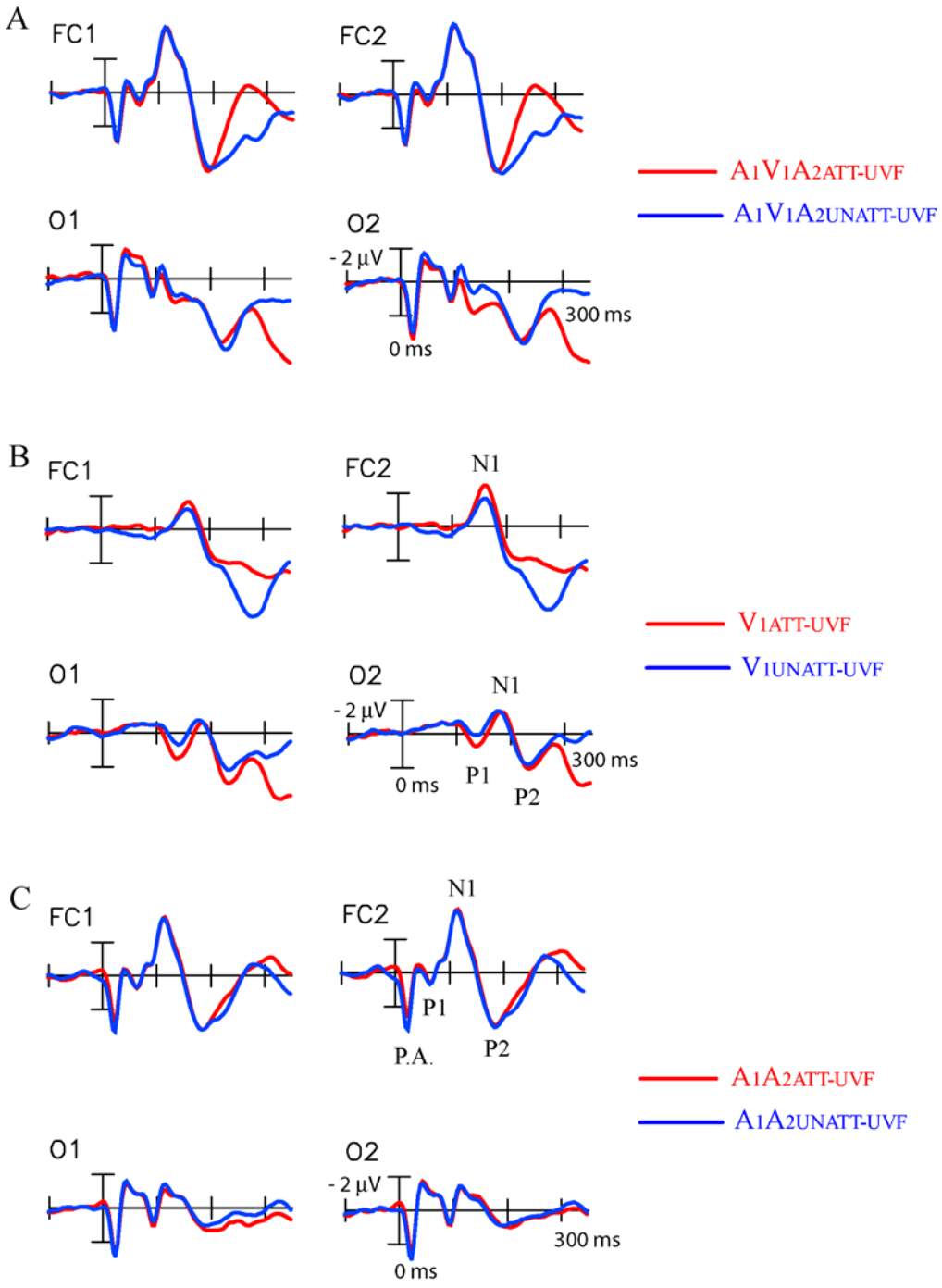

Figure 4.3 Grand-average ERP comparisons for the attended and unattended bimodal $A_1V_1A_2$ stimulus and its unimodal components V_1 and A_1A_2 presented in UVF. [A] ERPs elicited by the illusion-inducing $A_1V_1A_2$ stimulus when attended (ATT) and unattended (UNATT) [B] Corresponding ERPs as in [A] for the unimodal V_1 stimulus [C] Corresponding ERPs as in [A] for the unimodal A_1A_2 stimulus. Recordings are from left and right fronto-central (FC1,2) and occipital (O1,2) sites.

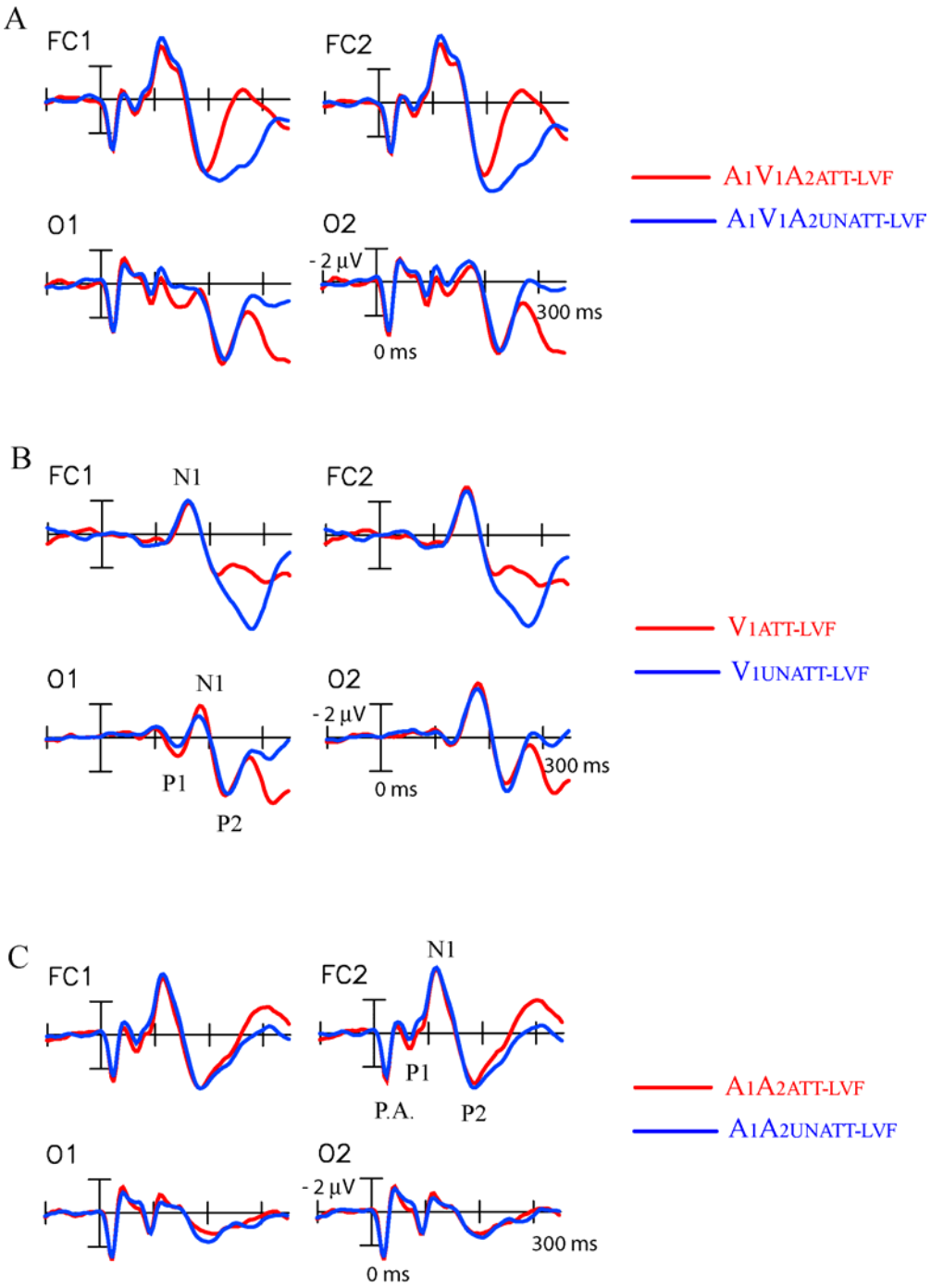

Figure 4.4 Grand-average ERP comparisons for the attended and unattended bimodal $A_1V_1A_2$ stimulus and its unimodal components V_1 and A_1A_2 presented in LVF. [A] ERPs elicited by the illusion-inducing $A_1V_1A_2$ stimulus when attended (ATT) and unattended (UNATT) [B] Corresponding ERPs as in [A] for the unimodal V_1 stimulus [C] Corresponding ERPs as in [A] for the unimodal A_1A_2 stimulus. Recordings are from left and right fronto-central (FC1,2) and occipital (O1,2) sites.

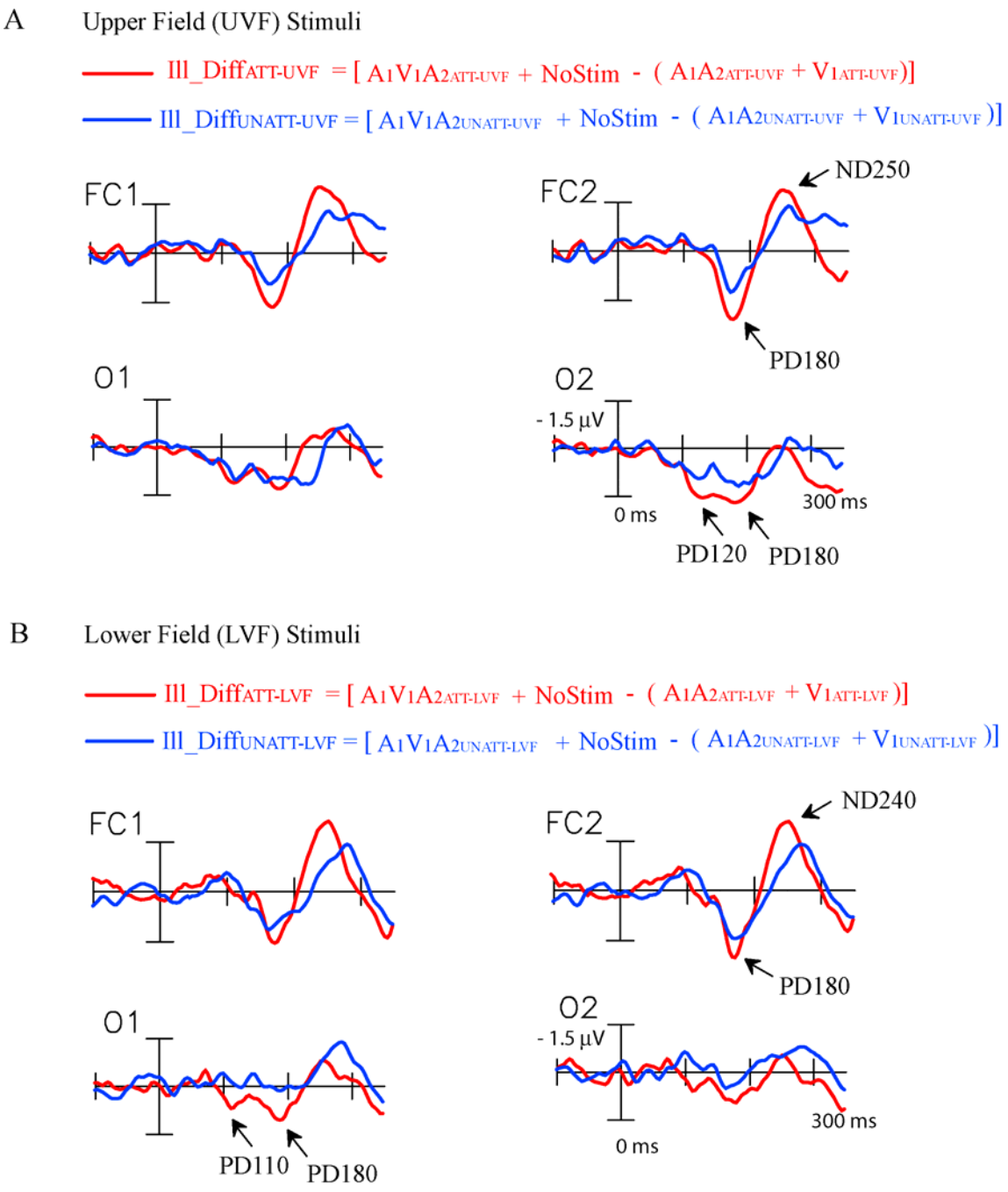

Figure 4.5 Grand-average Ill_Diff difference waves that reflect the cross-modal interactions within the illusion-inducing $A_1V_1A_2$ bimodal stimulus when attended (ATT) and unattended (UNATT). [A] Attended and unattended Ill_Diff difference waves in the upper visual field. [B] Corresponding difference waves as in [A] in the lower visual field. Recordings are from left and right fronto-central (FC1, 2) and occipital (O1,2) sites.

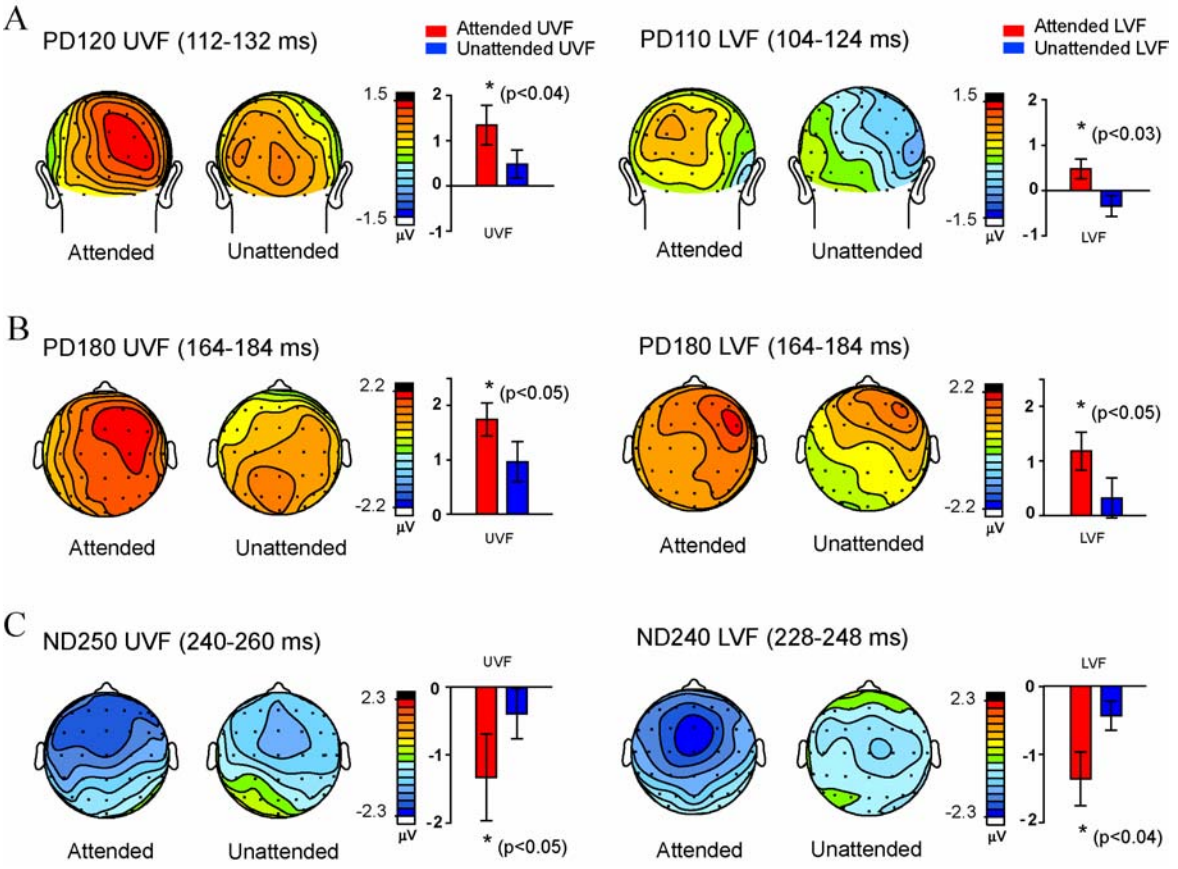

Figure 4.6 Topographical voltage maps of the three major components [A,B,C] in the Ill_Diff$_{ATT}$ and Ill_Diff$_{UNATT}$ difference waves for UVF in the left column and LVF in the right column shown in top or back views as per sites of maximal component amplitudes. Bar graphs next to the voltage maps depict the cumulative amplitude differences between the attended and unattended components.

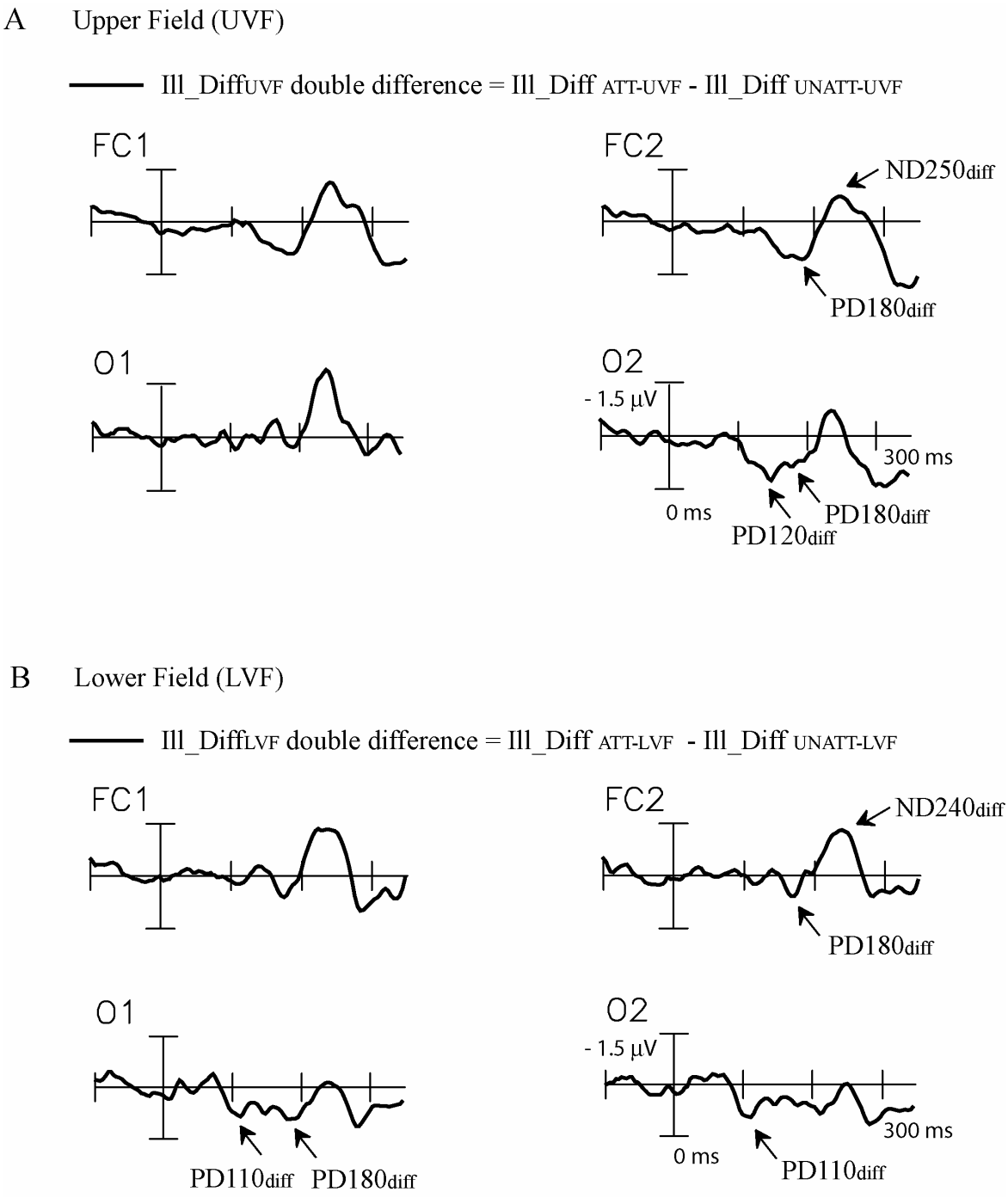

Figure 4.7 Grand-average Ill_Diff double difference waves that reflect the attention effects on the cross-modal interactions within the illusion-inducing $A_1V_1A_2$ bimodal stimulus [A] in the upper field (UVF) and [B] in the lower field (LVF). Recordings are from left and right fronto-central (FC1, 2) and occipital (O1,2) sites.

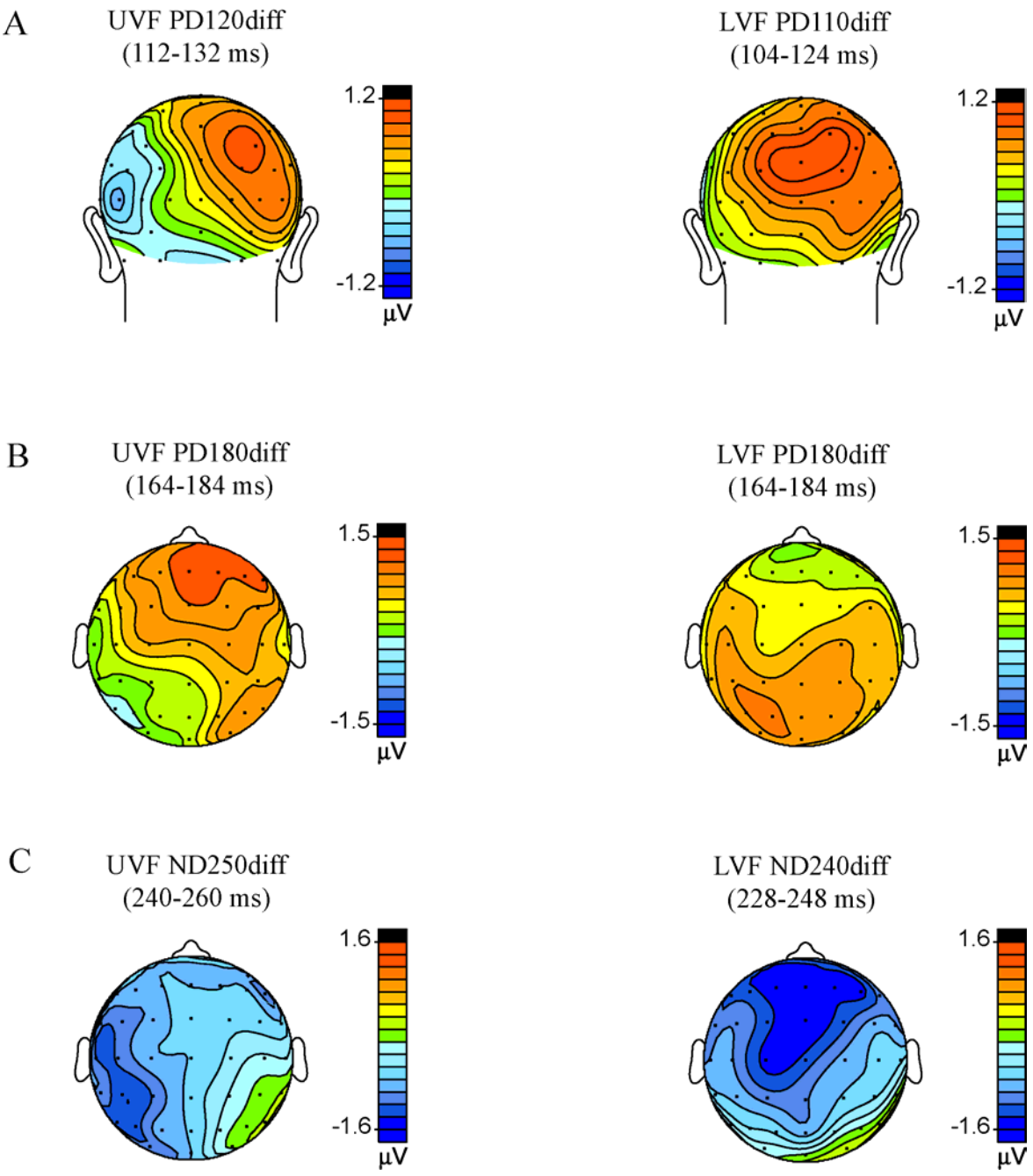

Figure 4.8 Topographical voltage maps of attention effects on the three major components [A,B,C] in the Ill_Diff_{UVF} and Ill_Diff_{LVF} difference waves in the left and right column respectively, shown in top or back views as per sites of maximal component amplitudes.

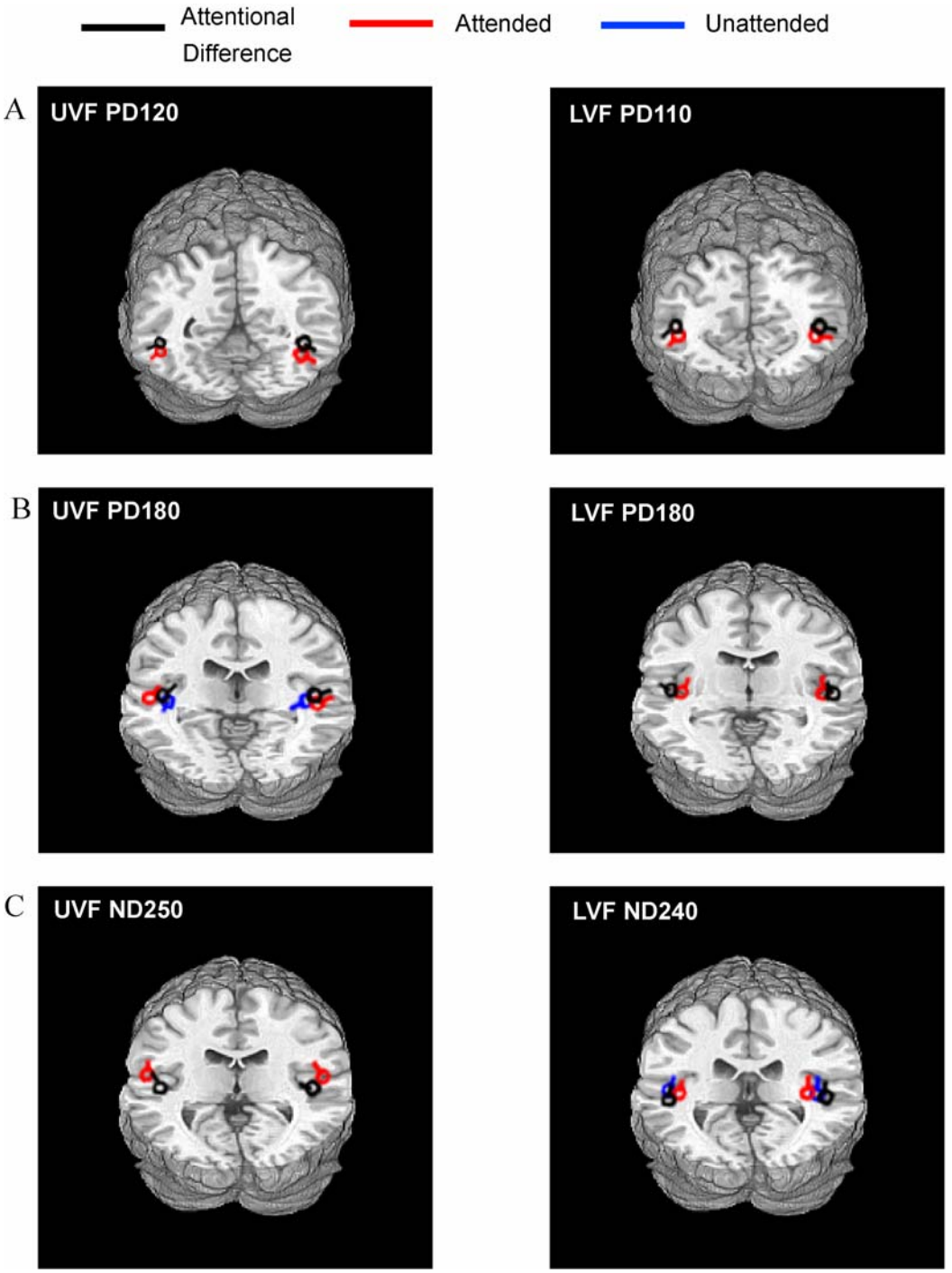

Figure 4.9 Estimated dipole sources modeled using BESA for the three sets of statistically significant components [A,B,C] in the grand-average Ill_Diff attended, unattended and attentional difference waveforms for UVF in the left column and LVF in the right column. Results are shown on a standard fMRI rendered brain in Talairach space.

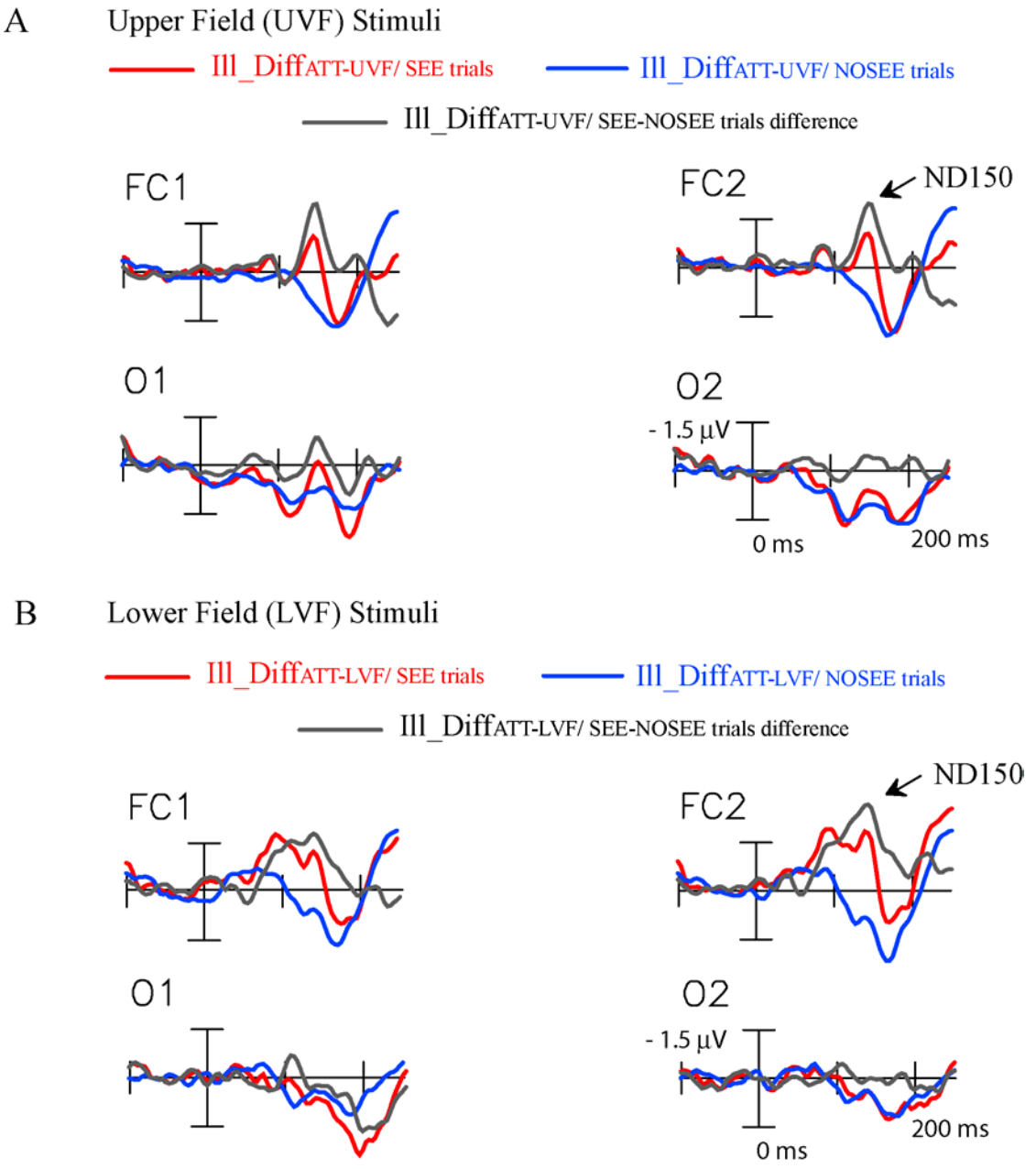

Figure 4.10 ERP differences between attended Ill_Diff trials on which the extra flash illusion was seen (SEE) vs. not seen (NO-SEE) [A] in the upper field (UVF) and [B] in the lower field (LVF). Recordings are from left and right fronto-central (FC1,2) and occipital (O1,2) sites.

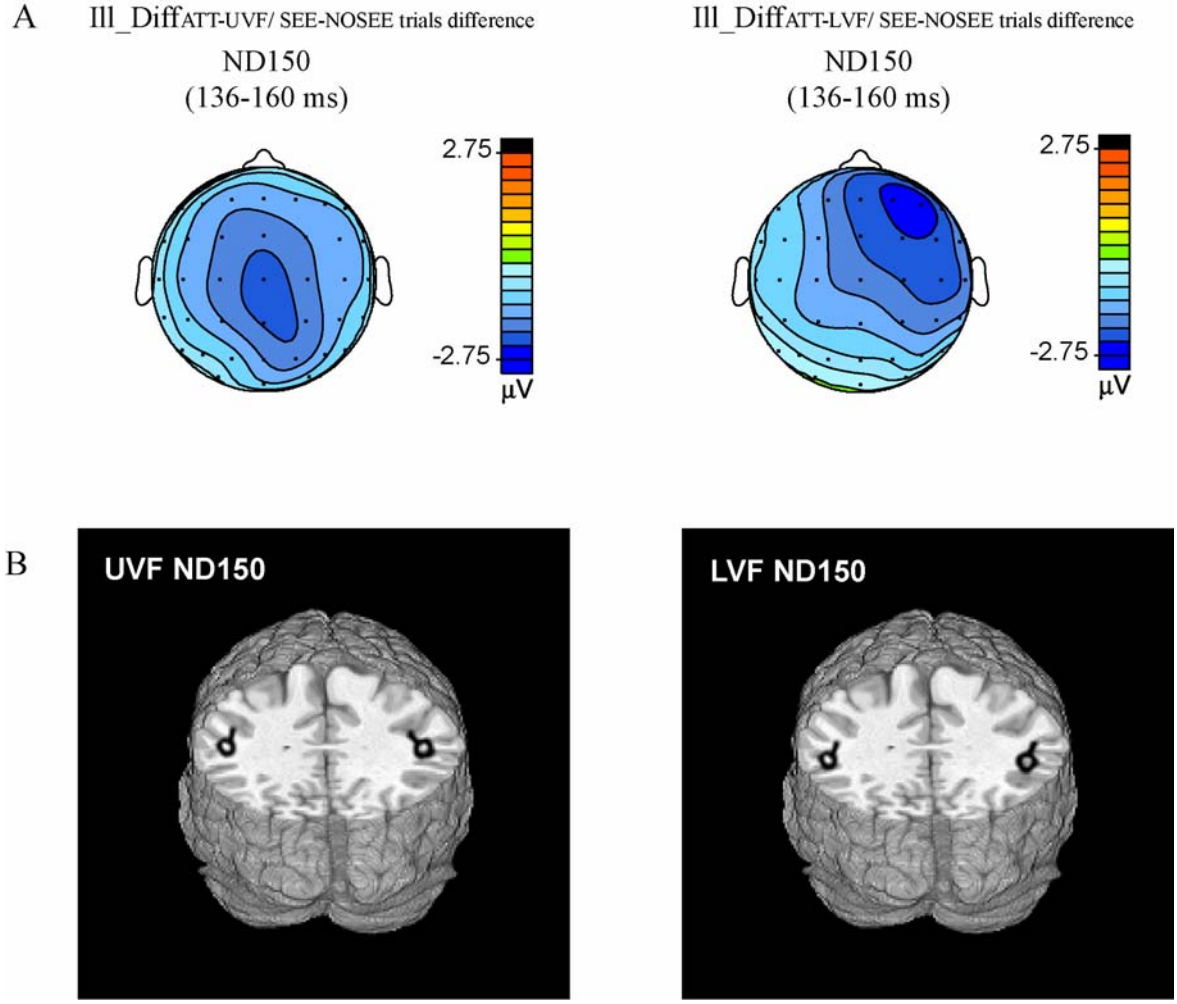

Figure 4.11 Topographical voltage maps and dipole sources of the main trial-specific component, ND150 in the SEE-NOSEE trials double difference wave shown for UVF in the left column and LVF in the right column.

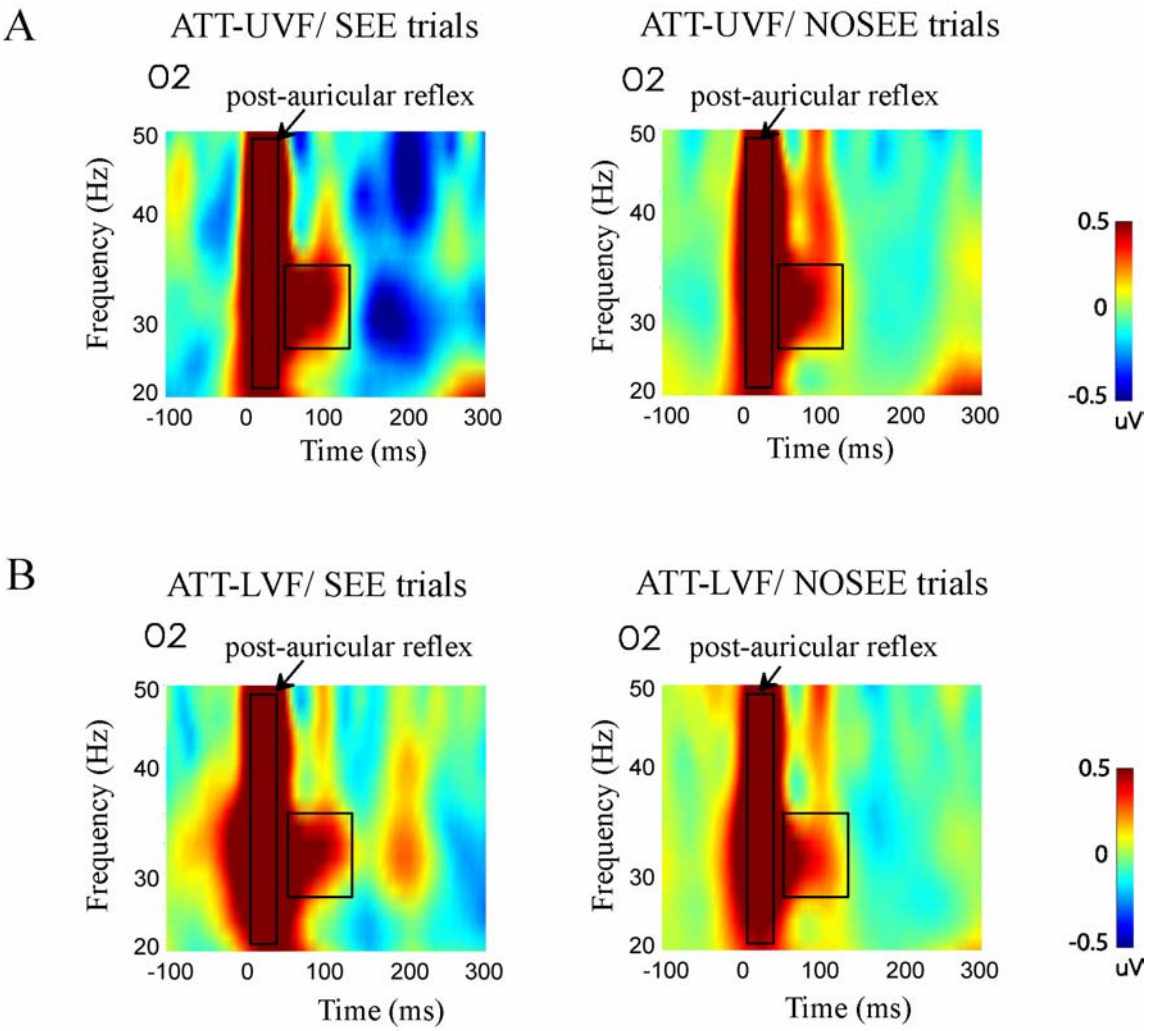

Figure 4.12 Frequency domain activity associated with perception of the sound induced illusory flash. [A] Time-frequency representation of the total average spectral amplitude on attended SEE trials and NO-SEE trials in UVF from an occipital electrode (O2). [B] Corresponding time-frequency representations as in [A] in LVF.

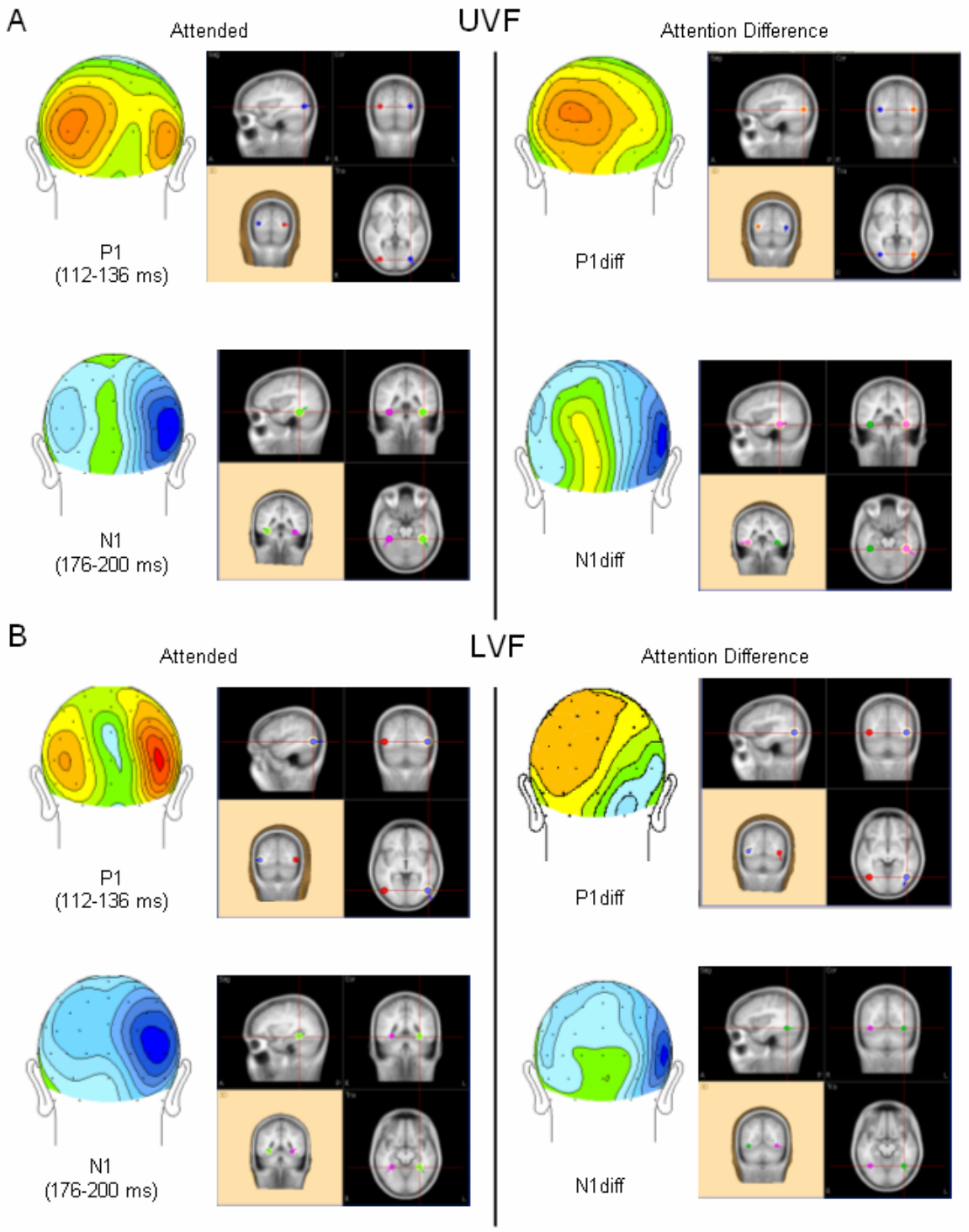

Figure 4.S1 Topographical voltage maps and BESA dipole fits for the P1 and N1 visual ERP components elicited by a single flash (V_1) in the attended ERP waveform in the left column and the attentional difference waveform in the right column in [A] the upper visual field and [B] lower visual field.

Table 4.1 Mean reaction times (SEM) for reporting the number of flashes seen (one or two) for all stimulus combinations containing one or two visual stimuli presented in the upper (UVF) and lower visual fields (LVF).

Stimulus	Mean RT (SEM) ms UVF [one / two flashes perceived trials]	Mean RT (SEM) ms LVF [one / two flashes perceived trials]
$A_1V_1A_2$	655 (13)	647 (11)
	[656(12) / 655(15)]	[654(9) / 639(15)]
V_1	654 (11)	656 (12)
V_1V_2	610 (9)	600 (10)
$A_1A_2V_1$	612 (14)	602 (16)

Table 4.2 Mean amplitudes of ERP components in the difference waves associated with the sound induced illusory flash generating $A_1V_1A_2$ stimulus when attended (ATT) or unattended (UNATT) in the upper (UVF) and lower (LVF) visual fields. Components were measured over scalp sites of maximal amplitude, as described in Methods. Significance levels of component amplitudes were tested with respect to the 100 ms pre-stimulus baseline.

	ERP Component	Amplitude (μV)	SEM (μV)	t(14)	p <
	PD120 (112-132 ms)	1.34	0.44	3.07	0.009
$Ill\_Diff_{ATT-UVF}$	PD180 (164-184 ms)	1.75	0.30	5.74	0.0001
	ND250 (240-260 ms)	-1.33	0.63	2.18	0.05
	(112-132 ms)	0.49	0.31	1.57	n.s.
$Ill\_Diff_{UNATT-UVF}$	PD180 (164-184 ms)	0.97	0.37	2.64	0.02
	(240-260 ms)	-0.39	0.37	1.05	n.s.
	PD110 (104-124 ms)	0.48	0.22	2.17	0.05
$Ill\_Diff_{ATT-LVF}$	PD180 (164-184 ms)	1.18	0.35	3.38	0.005
	ND240 (228-248 ms)	-1.36	0.40	3.43	0.005
	(104-124 ms)	-0.33	0.23	1.44	n.s.
$Ill\_Diff_{UNATT-LVF}$	(164-184ms)	0.32	0.37	0.88	n.s.
	ND240 (228-248 ms)	-0.47	0.21	2.29	0.04

Table 4.3 Mean amplitudes of ERP components in the attention double difference waves associated with the sound induced illusory flash in the UVF and LVF.

	ERP Component	Amplitude (μV)	SEM (μV)	t(14)	p <
Ill_Diff_{UVF}	$PD120_{diff}$ (112-132 ms)	0.85	0.36	2.33	0.04
double	$PD180_{diff}$ (164-184 ms)	0.77	0.34	2.24	0.05
difference	$ND250_{diff}$ (240-260 ms)	-0.96	0.35	2.19	0.05
Ill_Diff_{LVF}	$PD110_{diff}$ (104-124 ms)	0.81	0.32	2.56	0.03
double	$PD180_{diff}$ (164-184 ms)	0.86	0.39	2.22	0.05
difference	$ND240_{diff}$ (228-248 ms)	-0.94	0.39	2.39	0.04

Table 4.4 Talairach coordinates and corresponding brain regions of the dipole fits as modeled by BESA for the significant components in the attended and unattended Ill_Diff waveforms in UVF and LVF. Abbreviations are as, MTG: Medial Temporal Gyrus, STG: Superior Temporal Gyrus, vicinity implied ≤ 5 mm of proximity.

	ERP Component	x (mm)	y (mm)	z (mm)	Brain Region	R.V. (%)
$Ill\_Diff_{ATT-UVF}$	PD120	± 43	-55	-13	Inferior occipito-temporal cortex	8
	PD180	± 47	-30	1	MTG/ STG	9
	ND250	± 53	-21	11	vicinity of STG	6
$Ill\_Diff_{UNATT-UVF}$	PD180	± 41	-25	0	MTG/ STG	14
$Ill\_Diff_{ATT-LVF}$	PD110	± 42	-55	-1	Inferior occipito-temporal cortex	13
	PD180	± 40	-20	-1	MTG/ STG	11
	ND240	± 40	-29	4	vicinity of STG	6
$Ill\_Diff_{UNATT-LVF}$	ND240	± 42	-32	4	vicinity of STG	10

Table 4.5 Talairach coordinates and corresponding brain regions of the dipole fits as modeled by BESA for the significant components in the attentional difference Ill_Diff waveforms and the SEE-NOSEE trials Ill_Diff waveforms in UVF and LVF. Abbreviations are same as in table 4.4.

	ERP Component	x (mm)	y (mm)	z (mm)	Brain Region	R.V. (%)
Ill_Diff_{UVF} double difference	$PD120_{diff}$	± 40	-54	-11	Inferior occipito-temporal cortex	14
	$PD180_{diff}$	± 43	-26	-3	MTG/ STG	15
	$ND250_{diff}$	± 43	-21	3	vicinity of STG	18
Ill_Diff_{LVF} double difference	$PD110_{diff}$	± 46	-60	5	Inferior occipito-temporal cortex	17
	$PD180_{diff}$	± 44	-28	5	MTG/ STG	9
	$ND240_{diff}$	± 46	-33	2	vicinity of STG	7
$Ill\_Diff_{ATT-UVF/SEE-NOSEE}$	ND150	± 46	-5	4	STG	10
$Ill\_Diff_{ATT-LVF/SEE-NOSEE}$	ND150	± 49	-23	11	STG	8

Table 4.S1 Mean amplitudes of the P1 and N1 components elicited in ERPs to a single flash, V_1, when attended (ATT) or unattended (UNATT) in the upper (UVF) and lower (LVF) visual fields. Statistical differences between the attended and unattended component amplitudes are shown.

	ERP Component	ATT Amplitude (sem) μV	UNATT Amplitude (sem) μV	ATT vs. UNATT F(1,14), p<
V_{1UVF}	P1 (112-136 ms)	0.83 (0.37)	0.18 (0.26)	5.26, 0.04
	N1ant. (152-176 ms)	-1.85 (0.50)	-1.43 (0.35)	4.62, 0.05
	N1post. (176-200 ms)	-1.59 (0.40)	-1.05 (0.35)	4.71, 0.05
V_{1LVF}	P1 (112-136 ms)	0.85 (0.38)	0.58 (0.43)	5.13, 0.05
	N1ant. (152-176 ms)	-2.23 (0.45)	-2.06 (0.48)	2.64, n.s.
	N1post. (176-200 ms)	-2.01 (0.43)	-1.34 (0.44)	14.71, 0.002

REFERENCES

Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. Exp Brain Res 166:559-571.

Arden GB, Wolf JE, Messiter C (2003) Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. Vision Res 43:2469-2478.

Bertelson P (1999) Ventriloquism: A case of crossmodal perceptual grouping. In: Cognitive contributions to the perception of spatial and temporal events (Achersleben G, Bachman T, Musseler J, eds). Amsterdam: Elsevier.

Bertelson P, Vroomen J, de Gelder B, Driver J (2000) The ventriloquist effect does not depend on the direction of deliberate visual attention. Percept Psychophys 62:321-332.

Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA (2007) Neural basis of the ventriloquist illusion. Curr Biol 17:1697-1703.

Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci U S A 102:18751-18756.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110-1123.

Carrasco M (2006) Covert attention increases contrast sensitivity: Psychophysical, neurophysiological and neuroimaging studies. Prog Brain Res 154:33-70.

Clark VP, Fan S, Hillyard SA (1995) Identification of early visual evoked potential generators by retinotopic and topographic analyses. Hum Brain Mapp 2:170-187.

Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. Cogn Affect Behav Neurosci 4:117-126.

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci 3:292-297.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162-173.

Di Russo F, Martinez A, Sereno MI, Pitzalis S, Hillyard SA (2002) Cortical sources of the early components of the visual evoked potential. Hum Brain Mapp 15:95-111.

Di Russo F, Martinez A, Hillyard SA (2003) Source analysis of event-related cortical activity during visuo-spatial attention. Cereb Cortex 13:486-499.

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57:11-23.

Eimer M, Schroger E (1998) ERP effects of intermodal attention and cross-modal links in spatial attention. Psychophysiology 35:313-327.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749-5759.

Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Percept Psychophys 63:719-725.

Gondan M, Roder B (2006) A new method for detecting interactions between the senses in event-related potentials. Brain Res 1073-1074:389-397.

Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003) Visual localization ability influences cross-modal bias. J Cogn Neurosci 15:20-29.

Hillyard SA, Simpson GV, D.L. W, Van Voorhis S, Munte TF (1984) Event-related brain potentials and selective attention to different modalities. In: Cortical Integration (Renoso-Suarez F, Ajmone-Marsan C, eds), pp 395-414. New York: Raven Press.

Hillyard SA, Vogel EK, Luck SJ (1998) Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos Trans R Soc Lond B Biol Sci 353:1257-1270.

Hillyard SA, Anllo-Vento L (1998) Event-related brain potentials in the study of visual selective attention. Proc Natl Acad Sci U S A 95:781-787.

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3:284-291.

Hopfinger JB, Luck SJ, Hillyard SA (2004) Selective Attention: Electrophysiological and Neuromagnetic Studies. In: The Cognitive Neurosciences, III Edition (Gazzaniga M, ed), pp 561-574. Cambridge: MIT Press.

Jeffreys DA (1968) Separable components of human evoked responses to spatially patterned visual fields. Electroencephalogr Clin Neurophysiol 24:596.

Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282:108-111.

Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 23:315-341.

Kastner S, Pinsk MA (2004) Visual attention as a multilevel selection process. Cogn Affect Behav Neurosci 4:483-500.

Kiebel SJ, Tallon-Baudry C, Friston KJ (2005) Parametric analysis of oscillatory activity as measured with EEG/MEG. Hum Brain Mapp 26:170-177.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904-1911.

Luck SJ, Hillyard SA, Mouloua M, Woldorff MG, Clark VP, Hawkins HL (1994) Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection. J Exp Psychol Hum Percept Perform 20:887-904.

Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. Trends Neurosci 28:264-271.

Martinez A, DiRusso F, Anllo-Vento L, Sereno MI, Buxton RB, Hillyard SA (2001) Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. Vision Res 41:1437-1457.

Martinez A, Teder-Salejarvi W, Vazquez M, Molholm S, Foxe JJ, Javitt DC, Di Russo F, Worden MS, Hillyard SA (2006) Objects are highlighted by spatial attention. J Cogn Neurosci 18:298-310.

Martinez A, Teder-Salejarvi W, Hillyard SA (2007) Spatial attention facilitates selection of illusory objects: evidence from event-related brain potentials. Brain Res 1139:143-152.

McCarthy G, Wood CC (1985) Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol 62:203-208.

McDonald JJ, Teder-Salejarvi WA, Heraldez D, Hillyard SA (2001) Electrophysiological evidence for the "missing link" in crossmodal attention. Can J Exp Psychol 55:141-149.

McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2003) Neural substrates of perceptual enhancement by cross-modal spatial attention. J Cogn Neurosci 15:10-19.

McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2005) Neural basis of auditory-induced shifts in visual time-order perception. Nat Neurosci 8:1197-1202.

Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci 27:4120-4131.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115-128.

Molholm S, Martinez A, Shpaner M, Foxe JJ (2007) Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities. Eur J Neurosci 26:499-509.

Pick HL, Warren DH, J.C. H (1969) Sensory conflict in judgements of spatial direction. Percept Psychophys 6:203-205.

Picton TW, Hillyard SA, Krausz HI, Galambos R (1974) Human auditory evoked potentials. I. Evaluation of components. Electroencephalogr Clin Neurophysiol 36:179-190.

Posner MI, Petersen SE (1990) The attention system of the human brain. Annu Rev Neurosci 13:25-42.

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078-1093.

Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. Annu Rev Neurosci 27:611-647.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19-26.

Rorden C, Brett M (2000) Stereotaxic display of brain lesions. Behav Neurol 12:191-200.

Scherg M (1990) Fundamentals of dipole source analysis. Auditory evoked magnetic fields and electric potentials (Grandori F, Hoke M, Roman GL, eds):40-69.

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. Curr Opin Neurobiol 15:454-458.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308.

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

Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. Neuroreport 12:3849-3852.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147-152.

Shams L, Iwaki S, Chawla A, Bhattacharya J (2005) Early modulation of visual cortex by sound: an MEG study. Neurosci Lett 378:76-81.

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge, MA: MIT.

Stein BE, London R, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci 8:497-506.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Tallon-Baudry C, Bertrand O, Peronnet F, Pernier J (1998) Induced gamma-band activity during the delay of a visual short-term memory task in humans. J Neurosci 18:4244-4254.

Talsma D, Kok A (2002) Intermodal spatial attention differs between vision and audition: an event-related potential analysis. Psychophysiology 39:689-706.

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098-1114.

Talsma D, Kok A, Slagter HA, Cipriani G (2007a) Attentional orienting across the sensory modalities. Brain Cogn.

Talsma D, Doty TJ, Woldorff MG (2007b) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex 17:679-690.

Teder-Salejarvi WA, Munte TF, Sperlich F, Hillyard SA (1999) Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. Brain Res Cogn Brain Res 8:327-343.

Teder-Salejarvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106-114.

Teder-Salejarvi WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. J Cogn Neurosci 17:1396-1409.

Vroomen J, Bertelson P, de Gelder B (2001a) Directing spatial attention towards the illusory location of a ventriloquized sound. Acta Psychol (Amst) 108:21-33.

Vroomen J, Bertelson P, de Gelder B (2001b) The ventriloquist effect does not depend on the direction of automatic visual attention. Percept Psychophys 63:651-659.

Vroomen J, de Gelder B (2004) Perceptual effects of cross-modal stimulation: ventriloquism and the freezing phenomenon. In: The Handbook of Multisensory Processes (Calvert GA, Spence C, Stein BE, eds), pp 141-150. Cambridge, MA: MIT Press.

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. Neuroimage.

Chapter 5: The Sound-Induced Extra Flash Illusion influences Visual Features

ABSTRACT

An audio-visual stimulus sequence consisting of two sounds and a light flash presented within a <100 ms time window generates the visual percept of two flashes. This phenomenon has been described as the sound-induced extra flash illusion (Shams et al., 2000, 2002). The present study investigated the sensory properties of the illusory extra flash when the actually presented flash had visual attributes such as color and basic shape. The visual properties of the illusory flash were also probed in the context of association of two sequential real flashes, wherein the first and second flashed stimuli possessed different visual attributes (such as different colors or differently oriented line shapes), which could fuse to form a flash with a blend of the attributes (such as a new color or shape). It was found that the illusory extra flash had the same feature properties as the actual flash, both for simple features of color and shape and for the conjunction of these features. The extra flash illusion was also seen with the fused percept resulting from the merging of sequential visual stimuli. Apart from the illusory flash phenomenon, however, the perceptual fusion of two sequential visual features was actually reduced in the presence of two pulsed sounds. These results suggest that temporal properties of visual stimuli can be influenced by audition in a variety of ways.

INTRODUCTION

When a single flash of light is presented interposed between two brief auditory stimuli separated by 60-100 ms, subjects typically report perceiving two flashes, of which the second is illusory (Shams et al., 2000, 2002). The neural basis of the extra illusory flash percept has been investigated in detail in electrophysiological and neuroimaging studies (Shams et al., 2001, 2005, Arden et al., 2003, Watkins et al., 2006, Mishra et al., 2007). In a detailed investigation of the phenomenon using event related potential recordings (ERPs), Mishra et al. (2007) found that the illusion is based on a rapid interplay between auditory, visual as well as polysensory areas that occurs within 150 ms of stimulus onset. Early visual cortex activation was observed in association with the sound induced illusory flash, but the neural activity pattern underlying this phenomenon was found to be very different from the activity elicited by a real second flash in a double flash stimulus. These differences raise the question as to what is the perceptual nature of the illusory flash and whether it can have specific attributes akin to real visual stimuli such as contrast, color, shape and orientation.

A recent psychophysical study (McCormick and Mamassian, 2007) has suggested that the illusory flash can have a measurable contrast. The investigators used a paradigm wherein the first real flash within the illusion-generating stimulus was presented at two spatial locations, in the left as well as the right hemifield. The first flash was either high contrast white or black and was followed by a second flash of threshold-level contrast at only one of the two locations. The contrast threshold of the second flash was compared in the presence of either one or two tones concurrent to the visual stimulation. It was found that in the case of high contrast white flashes the second flash indeed had a lower contrast

at threshold in the presence of two tones relative to one tone. Moreover, this difference existed only when the second flash was the same contrast polarity as the first white flash. It was concluded that the facilitation of visibility of the threshold second flash in the presence of two tones was a result of its interaction with the spatio-temporally concurrent illusory flash. Hence, under certain stimulus conditions the sound-induced illusory flash was shown to have psychophysically assayable characteristics such as contrast similar to real flashes.

In the present study, we investigated the perceptual nature of the sound-induced extra flash illusion by studying specific visual attributes that the illusion may affect such as color and simple shapes in the form of oriented lines, as well as simple objects formed by a spatial conjunction of these two features. In addition, we investigated whether the sound-induced visual illusion can affect the temporal fusion of visual features presented in rapid sequence. This followed a similar logic as the experiment by McCormick and Mamassian (2007), namely that the illusory flash induced by the first flash and two pulsed sounds may interact with a second real flash and alter the temporal fusion between the two real flashes. However, depending on the visual processing stage and latency at which audition alters visual perception, it is also possible that the pulsed sounds may not affect the process of temporal fusion of visual features per se, but instead, may influence the temporal properties of the final fused percept. The present experiment distinguishes between these possibilities by monitoring perceptual reports of the observed visual features in conjunction with illusory flash reports.

MATERIALS AND METHODS

Task and Stimuli

Seventeen right-handed healthy adults (8 males and 9 females, mean age 23.4) participated in the study after giving informed consent. Each participant had normal or corrected-to-normal vision and normal hearing. Participants in the study were pre-selected as individuals who perceived the sound-induced extra flash illusion on more than half the trials in a short 2 min. screen prior to experimentation.

The experiment was conducted in a sound-attenuated chamber having a background sound level of 32 dB and a background luminance of 2 cd/m^2. Visual stimuli were presented at 8º of visual angle (va) in the left visual field on an LCD monitor at a 60 Hz refresh rate in a darkened room. Subjects maintained fixation on a cross positioned at the centre of the mid-level grey screen at a viewing distance of 83 cm. Auditory stimuli were presented from speakers attached one on each side of the monitor display. Auditory stimuli were 10 ms 76 dB noise bursts. Three types of visual stimuli were used in Experiment 1 representing the visual attributes of i) color ii) shape and iii) the conjunction of color and shape. Visual stimuli were flashed for 32 ms. Each visual attribute was tested in a separate block. In experiment 1 the basic illusion was tested on the various visual attributes. Subjects were presented with 12 trials each of five randomized visual or audio-visual stimuli, V_1, V_1V_2, A_1V_1, $A_1V_1A_2$ and $A_1A_2V_1$, and reported perception of either one or two visual stimuli on each trial. In this terminology, suffixes 1 or 2 denote the first or second occurrence of the auditory, A, or visual, V, component of each stimulus combination. The SOA between A_1A_2 was 67 ms, V_1V_2 84

ms and A_1V_1 3 ms in every stimulus combination that included them. In $A_1A_2V_1$, V_1 followed A_1 by 200 ms, and this stimulus served as a stimulus-matched behavioral control for the $A_1V_1A_2$ test stimulus (Mishra et al., 2007). Any subjects who failed to perform at >90% accuracy on the control stimulus were excluded from the experiment. All subjects included here met this criterion.

The visual stimuli in the color block were orange annuli having an outer diameter and thickness of 3.7º and 0.8º va, respectively. In the shape block the visual stimulus was an achromatic symmetric cross formed by the juxtaposition of a horizontal and a vertical line each made of a single cycle of a 2 cycles/degree gabor. The length/width and thickness of the cross were 3.7º and 0.8º va, respectively, similar to the dimensions of the annulus on the color block. In the feature conjunction block the achromatic cross was replaced by an orange cross. The stimuli in experiment 1 (except the $A_1A_2V_1$ control condition) are shown in Fig. 1A. In all conditions subjects were instructed to report seeing either one or two flashes. Importantly, two flashes were to be reported only when they saw the same stimulus (orange annulus, grey cross, or orange cross) flash twice.

The above stimuli were specifically chosen because each could be constructed by the juxtaposition of simple visual features presented in rapid succession. For example, perception of an orange annulus is reported when presentation of a red annulus is rapidly followed by a spatially congruent green annulus. Similarly the (achromatic/ orange) cross is perceived by rapid presentation of a (achromatic/ orange) vertical line following a (achromatic/ orange) horizontal line (or vice versa). Annuli were chosen in the color block instead of simple disks as sequential color fusion was found to be more efficient

using annuli than disks. The goal of experiment 2 was to investigate the influence of the pulsed sounds on the sequential fusion of visual features.

Experiment 2 was carried out in three separate blocks for i) color ii) shape and iii) the conjunction of color and shape. Sequential fusion of visual features was studied at four V_1V_2 SOAs of 50 ms, 84 ms, 100 ms and 116 ms. The second visual stimulus was presented as a 16 ms flash to optimize its temporal fusion with the first flash. In each block, 12 trials of each visual stimulus combination at each SOA were presented in conjunction with two pulsed sounds (67 ms SOA), one sound or no sound, designated as $A_1V_1A_2V_2$, $A_1V_1V_2$, and V_1V_2, respectively. An equal number of control stimuli wherein the second flash was absent ($A_1V_1A_2$ and A_1V_1 controls) or was the same as the first flash (V_1V_2 control) were also presented. All stimuli were equiprobable and presented in random order. The stimulus design for experiment 2 is shown in Fig. 1B

Within the color block subjects reported perception of every visual stimulus by making one of four color choices using one of four buttons on a joystick: i) red ii) green iii) orange or iv) both red and green perceived separately. Choice iii) orange was perceived on those trials when the red and green annuli presented sequentially merged to a single orange annulus. Along with a color report, subjects also reported perception of the sound induced extra flash illusion by making a separate button press only if the annulus having the perceived color (red, green, or orange) was seen to flash twice. For the color choice iv) on which both red and green stimuli were perceived as distinctly apart, the separate button response for two flashes was made only if either the red or the green component was seen to flash twice.

In the shape as well as color-shape conjunction blocks subjects gave perceptual reports analogous to the color block. Subjects made one of four button choices to indicate the perceived shape of each visual stimulus (i) horizontal line (ii) vertical line (iii) cross formed by the temporal merge of the horizontal and vertical lines or (iv) both horizontal and vertical lines seen separately. Parallel to the color block, subjects used a separate button to indicate if the shape perceived on each trial was seen to flash twice. Again, for shape choice iv) on which horizontal and vertical lines were perceived apart, the separate button response for two flashes was made only if either the horizontal or vertical line was seen to flash twice. The task on the color-shape conjunction block was identical to the shape block except the stimuli were constructed with orange colored horizontal and vertical lines instead of achromatic lines.

In Experiment 3 the influence of pulsed sounds on sequential fusion of visual features was analyzed by using variable auditory (A_1A_2) SOAs, while keeping the visual (V_1V_2) SOA constant at an intermediate SOA of 84 ms. Three auditory SOAs were used: 33 ms, 67 ms and 99 ms. Visual feature perception was tested in a color block and an achromatic shape block with the same visual stimuli and response choices as in Experiment 2. The stimulus types used in Experiment 3 are shown in Fig. 1C.

RESULTS

Experiment 1

Figure 2 shows that for all three types of visual stimuli tested (orange annulus, achromatic cross and orange cross), subjects reported perceiving the $A_1V_1A_2$ stimulus combination as containing two flashed stimuli on a significant proportion of trials,

averaging 46% overall. There was no significant difference between percentage of illusory reports between the three types of visual stimuli, indicating that the two pulsed sounds equivalently affected all three ($F(2,32) = 0.28$, p = n.s.). Perceptual reports of single or double flashes for the other stimulus combinations (A_1V_1, V_1 and V_1V_2) were highly accurate on the color, shape and color-shape conjunction blocks (Fig. 2).

Experiment 2

The effect of pulsed sounds on the temporal fusion of sequential visual stimuli is represented in Fig. 3 and 4. The perceptual fusion of visual features was maximal at the shortest flash SOA of 50 ms and showed a steady decrement at every subsequent SOA. This trend was seen in case of both visual features investigated, the fusion of red and green to an orange annulus in case of color, and the merging of oriented horizontal and vertical lines to form a cross in case of shape. The same result was also observed for colored cross stimuli that contained both color and shape attributes. The data in Fig. 3 is alternately plotted in Fig. 4 to highlight the decrement in feature fusion with increasing visual SOA. The effect of SOA on fusion over all stimulus combinations was highly significant (color: $F(3,48) = 146.31$, $p < 0.0001$; achromatic shape: $F(3,48) = 157.51$, $p < 0.0001$; colored shape: $F(3,48) = 50.02$, $p < 0.0001$). A significant effect of stimulus condition ($A_1V_1A_2V_2$ / $A_1V_1V_2$ / V_1V_2) on temporal fusion was found for color ($F(2,32) = 4.25$, $p < 0.03$) and achromatic shape ($F(2,32) = 8.99$, $p < 0.0008$) but not for the colored shape ($F(2,32) = 0.87$, p = n.s.). The ISI x stimulus condition interaction was found to be significant for color ($F(6,96) = 3.90$, $p < 0.002$) and achromatic shape ($F(6,96) = 4.10$, $p < 0.002$), but it did not reach significance for the colored shape ($F(2,32) = 2.08$, p = n.s.).

Post-hoc analysis showed that for both the colored annuli and the achromatic cross stimuli sequential fusion between the $A_1V_1A_2V_2$, $A_1V_1V_2$ and V_1V_2 combinations differed mainly at the 84 ms visual SOA. At this SOA, color fusion in the presence of two sounds as in $A_1V_1A_2V_2$, was significantly diminished relative to the other two stimulus combinations ($F(2,32) = 10.85$, $p< 0.0003$) (Fig. 4A). At the same SOA, the merging of achromatic lines to form a cross was also significantly reduced to the two sound (and no sound) conditions in relation to the one sound condition (Fig. 4B) ($F(2,32) = 10.81$, $p< 0.0003$).

Perceptual two flash reports in Experiment 2 are shown in Fig. 5. In presence of the double sound bursts, the color (orange) or shape (achromatic/ colored cross) resulting from the fusion of the first and second visual stimulus was perceived to flash twice significantly more frequently than in the presence of one or no sounds. The extra flash illusion was significant when tested over all visual SOAs and over both types of perceived trials; i.e., both fused and unfused percept trials (flash illusion compared between $A_1V_1A_2V_2$, $A_1V_1V_2$, V_1V_2: color: $F(2,32) = 28.97$, $p< 0.0001$; achromatic shape: $F(2,32) = 12.77$, $p< 0.0001$, colored shape: $F(2,32) = 10.17$, $p< 0.0004$). Two flash responses at the longer visual SOAs of 100 ms and 116 ms are not plotted in Fig. 5 as they only occurred on a small percentage of $A_1V_1A_2V_2$ trials (6% and 25% of fused and unfused percepts, respectively, over all stimulus types). Perceptual responses to the control stimuli (Fig. 1B) in experiment 2 were similar to responses to identical stimuli in experiment 1 and hence are not shown.

Individual subject data for the 50 ms visual SOA are plotted in Fig. 6 showing the relative occurrence of two flash reports for the $A_1V_1A_2V_2$ stimuli compared to $A_1V_1V_2$

stimuli. Data on both axes are normalized with respect to two flash reports on the V_1V_2 stimulus which could be considered erroneous responses as V_1V_2 was reported as a fused color/shape flash at this short SOA (Fig. 3 and 4). The majority of the data points lay in the upper left quadrant indicating that presence of two sounds facilitated two flash generation compared to a single sound. This occurred for the fused percept resulting from the sequential merge of colors as well as achromatic and colored shapes. Data point clusters for the three visual attributes tested could not be separated from one another, indicating that the influence of sounds was the same for each attribute (color vs. shape vs. colored shape two flash reports: $F(2,32) = 1.25$, p = n.s.).

Experiment 3

The SOA between auditory stimuli was varied within experiment 3 at a constant visual SOA of 84 ms, to assess if the timing of the paired sounds affects temporal association between sequential visual stimuli. Flashes were either red and green annuli fusing to orange in the color block, or achromatic horizontal and vertical lines forming a cross in the shape block. The auditory SOAs were varied at 33 ms, 67 ms and 99 ms. For both types of stimuli, the presence of two pulsed sounds did not produce increased fusion of the first and second visual stimulus relative to co-occurrence of a single sound or no sound (Fig. 7). In fact, color fusion (i.e., proportion of 'orange' reports) was markedly diminished at the 33 ms and 99 ms auditory SOAs and line merging (i.e., proportion of 'cross' reports) at 66 and 99 ms SOAs within $A_1V_1A_2V_2$ relative to the one or no sound stimulus combinations (color fusion across $A_1V_1A_2V_2$/ $A_1V_1V_2$/ V_1V_2: 33 ms ($F(2, 32) = 5.75$, $p < 0.008$), 66 ms ($F(2, 32) = 2.61$, p = n.s.), 99 ms ($F(2, 32) = 6.84$, $p < 0.002$); line

overlap: 33 ms ($F(2, 32) = 2.42$, p = n.s.), 66 ms ($F(2, 32) = 8.95$, $p< 0.0009$), 99 ms ($F(2, 32) = 15.39$, $p< 0.0001$)).

Consistent with experiment 2, the color or shape percept was found to be influenced by the sound induced-extra flash illusion (Fig. 8). At both 33 ms and 66 ms auditory SOAs the proportion of trials on which double flashes were reported for either perceived color or shape was significantly greater on $A_1V_1A_2V_2$ versus $A_1V_1V_2$/ V_1V_2 trials (extra flash illusion: color: 33 ms ($F(2, 32) = 9.73$, $p< 0.0006$), 66 ms ($F(2, 32) = 8.01$, $p< 0.002$), 99 ms ($F(2, 32) = 2.77$, p = n.s.); shape: 33 ms ($F(2, 32) = 4.88$, $p< 0.02$), 66 ms ($F(2, 32) = 4.49$, $p< 0.02$), 99 ms ($F(2, 32) = 2.02$, p = n.s.)).

The $A_1V_1A_2$ stimulus, where V_1 was an orange annulus or an achromatic cross on color and shape blocks, respectively, was also tested at the different auditory SOAs. The extra flash illusion reports were significantly greater on these $A_1V_1A_2$ trials, relative to $A_1V_1A_2V_2$ trials for every auditory SOA for color as well as shape (Fig. 8) (extra flash illusion, $A_1V_1A_2$ vs. $A_1V_1A_2V_2$: color: ($F(1, 16) = 20.49$, $p< 0.0003$), shape: ($F(1, 16) = 69.21$, $p< 0.0001$)), suggesting that presence of V_2 may actually be interfering with the illusion phenomenon.

DISCUSSION

In the present study the sound-induced extra flash illusion was explored in the context of specific visual attributes such as color and basic line shapes. In Experiment 1 it was found that both colored stimuli (orange annuli) and stimuli with basic shapes (achromatic and colored line crosses) were susceptible to the extra flash illusion. To investigate whether presence of sounds can qualitatively alter visual feature perception,

the sound induced flash illusion was studied in the context of fusion of sequentially presented visual features, for both colors and line shapes. This experiment was performed at different visual SOAs (experiment 2) and consecutively at variable auditory SOAs (experiment 3). The initial hypothesis was that in the presence of two sounds, the temporal fusion between the first and second flashed stimulus may be qualitatively enhanced due to the co-occurrence of the extra flash illusion to the first real flash at the time of the second flash. Temporal fusion was indeed found to be altered by two sounds, but contrary to the initial hypothesis, temporal fusion between sequential visual stimuli was actually reduced on two sound trials relative to presence of one or no sounds. It was also found that the fused percept resulting from the sequential presentation of two flashes at short SOAs was modulated by the presence of sounds. In particular, an orange stimulus perceived as a result of fusion of sequential red and green flashed stimuli, was seen to flash twice in the presence of two sounds. Similarly, a cross formed by the overlap of sequentially presented horizontal and vertical lines was perceived to flash twice in the presence of two sounds. Hence, the extra flash illusion was observed for all visual attributes, but as a distinct process from sensory fusion between the visual components of the stimulus.

In the first experiment, it was demonstrated that the extra flash illusion generates a perceptual replication of the visual stimulus presented between two sounds; this was true for all visual objects tested that had color or basic shape features. Within Experiments 2 and 3, the temporal fusion of sequential colored stimuli (red and green merging to form orange) as well as line shapes (horizontal and vertical lines merging to form a cross) was primarily observed at short first-second flash SOAs and short first-

second sound SOAs, respectively. In both experiments, the extra flash illusion was consistently observed on the fused percept (as well as on unfused percept trials). In neither case, however, was the quality of the fused feature enhanced by the presence of two sounds (i.e., more orange/ cross perception). This could have occurred if the illusory flash was generated to the first flashed stimulus and shared its visual properties such that presence of the illusion facilitated the fusion with the subsequent real flash. Rather it was found that two sounds tended to reduce temporal fusion, as if the two real flashes were kept separate in correspondence with two sounds.

The finding that the temporal properties of the final percept are modified by the sounds is in line with the 'modality appropriateness hypothesis' (Welch and Warren, 1980, Welch et al., 1986). Audition dominates the temporal component (number of flashes) of the final percept as temporal coding is most accurate within audition, while vision dominates the form (color/ shape) of the percept as it has better resolution within this domain. Some facilitation by sounds on visual properties such as contrast have been reported, but only when a sound provided information about the visual stimulus such as the timing of the display (Lippert et al., 2007). The previous investigation on the visual properties of the illusory flash (McCormick and Mamassian, 2007) also did not find consistent effects with respect to stimulus contrast using two different stimulus types with opposite contrast polarity. In that study, however, the illusory flash was inferred to be a true percept on the basis of a signal detection analysis. Overall, the illusory flash appears to be a significant perceptual phenomenon but because it is generated in a unique audio-visual context it does not seem to have the exact visual properties as a real flash.

The reduction in temporal fusion of sequential visual stimuli in presence of two sounds relative to one or no sound, which was most significant at the intermediate visual SOA of 84 ms, is a novel finding. Although this findings was unexpected with regards to the initial hypothesis based on the extra flash illusion discussed above, it is consistent with audio-visual matching of the number of sounds with the number of flashes. Within the $A_1V_1A_2V_2$ stimulus configuration, the first and second sound may temporally bind to the first and second flash, respectively, and effectively separate the two flashes. Moreover, results from experiment 3 suggest that matching of the number of auditory and visual components of a stimulus may even interfere with and reduce the extra flash illusion phenomenon; that is, reports of the illusory extra flash were much reduced on $A_1V_1A_2V_2$ trials compared to $A_1V_1A_2$ trials. Synchronization of the auditory and visual components of audio-visual stimuli is a robust phenomenon, and the flash rates of visual stimuli have been shown to be drawn in temporal alignment with auditory events, such that a single temporal marker exists for the total multisensory combination (Morein-Zamir et al., 2003, Berger et al., 2003, Recanzone et al., 2003). Our findings are thus in line with these reports and emphasize the importance of audio-visual matching, which is most crucial to forming coherent percepts of natural audio-visual stimuli (Stein and Meredith, 1993, King and Calvert, 2001, Spence and Squire, 2003, Burr and Alais, 2006).

Finally, the present study sheds some light on the processing time window of the extra flash illusion. For the short flash SOAs of 50 ms and 84 ms at which fused percepts were seen on a significant number of trials for both colors and oriented lines, the extra flash illusion was consistently reported for the final fused percept. This suggests that the sound-induced visual illusion may affect visual processing at a longer latency than visual

sensory fusion. Association between sequential colors can occur rapidly, even within the retina and the lateral geniculate nucleus before visual information reaches visual cortex. However, the association of oriented lines to form a cross is most likely a cortical process that follows stimulus encoding within the orientation columns in primary visual cortex. That both the resulting color (orange) and shape (cross) formed by the fusion of activity to the first and second flash were equally affected by the sound-induced illusion suggests that sounds may affect visual processing at later points in the visual hierarchy than primary visual areas.

In sum, the present study provides additional insight into the sound-induced extra flash illusion phenomenon and demonstrates that the illusion can incorporate specific visual attributes such as color and basic shapes. Apart from the illusory flash phenomenon, it was also found that presence of two sounds tends to reduce the sensory fusion between sequential visual stimuli relative to one or no sounds. Thus, audition can alter the temporal properties of visual stimuli in more ways than one.

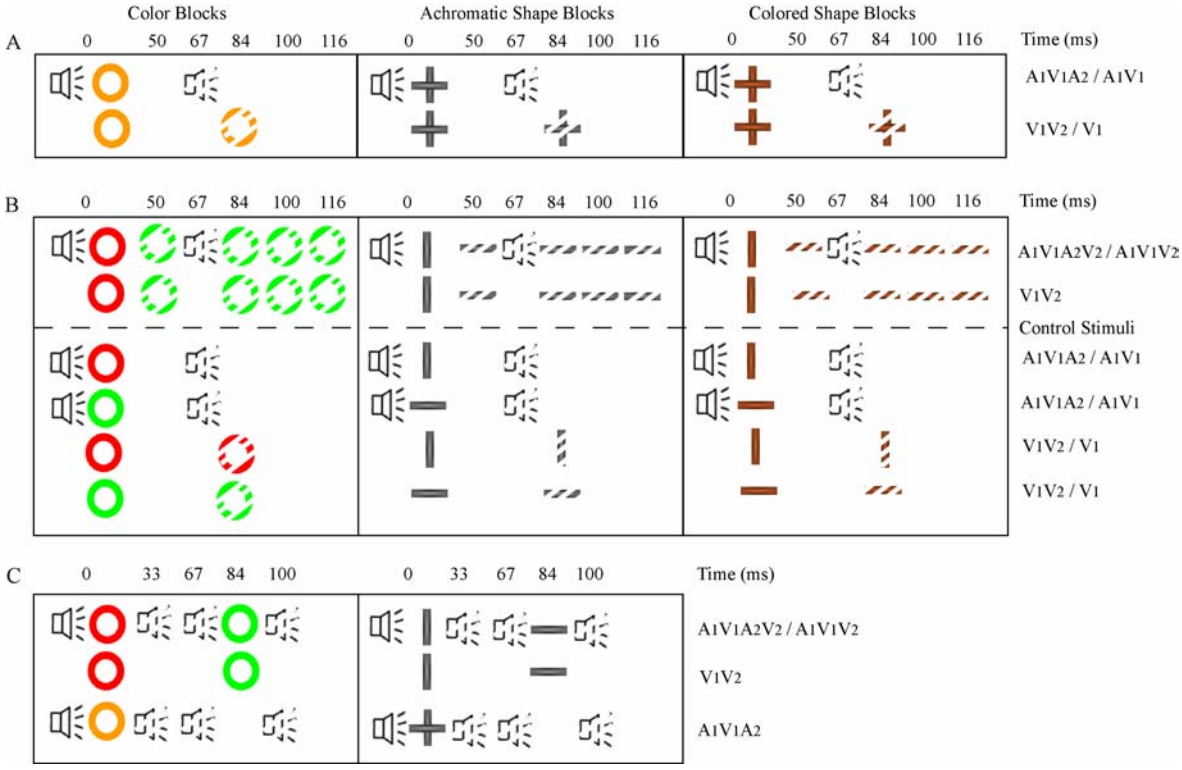

Figure 5.1 Overview of the experimental design in [A] Experiment 1 [B] Experiment 2 and [C] Experiment 3. Each experiment was performed in separate blocks for color, achromatic shape and colored shapes. Stimuli drawn in dashed lines indicate the presence or absence of that stimulus in different audio-visual configurations or at different onset asynchronies (SOAs). Auditory (A) and visual (V) stimuli are labeled 1 or 2 to designate their first or second occurrence in each stimulus configuration.

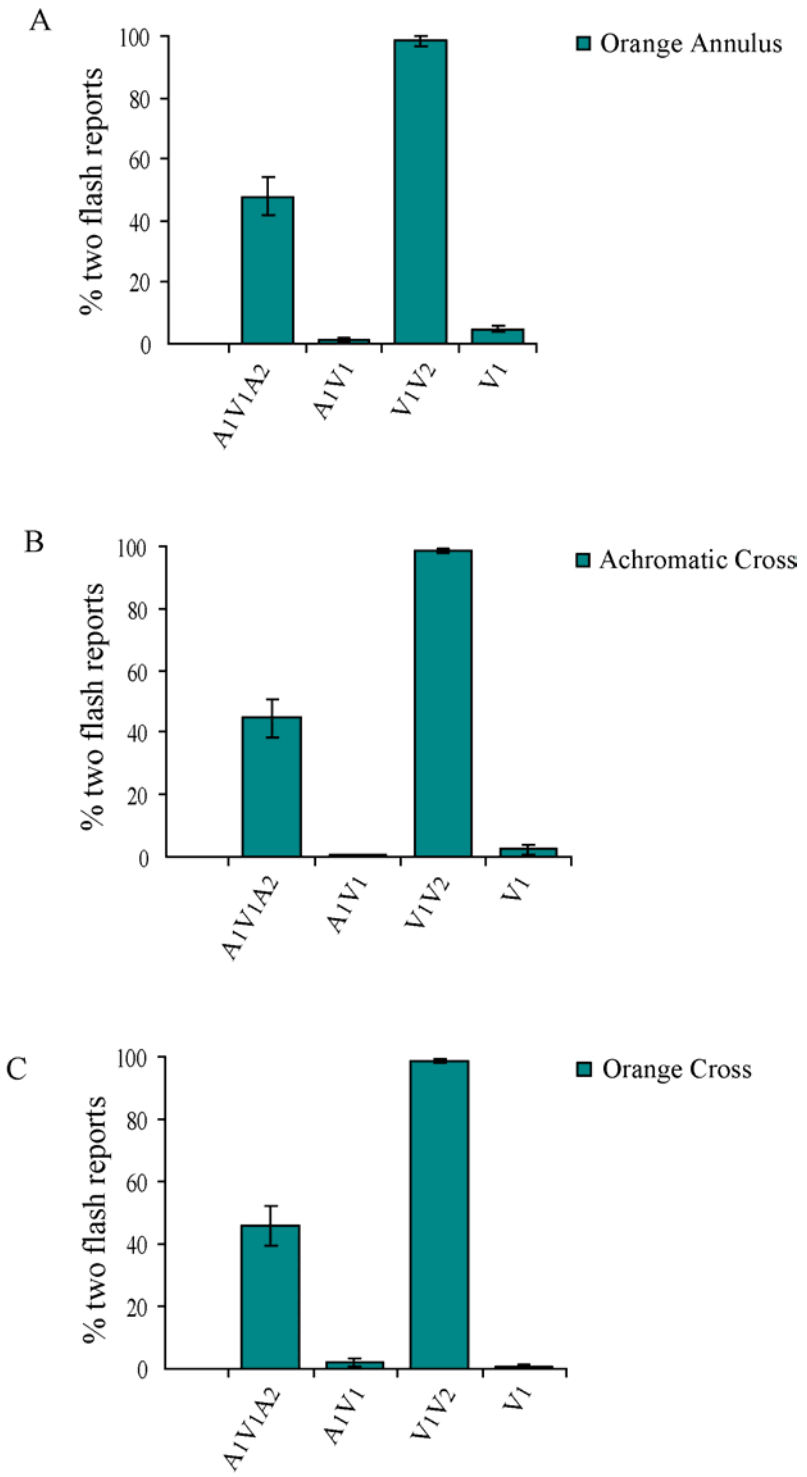

Figure 5.2 Percent two flash reports for audio-visual/ visual stimuli tested in Experiment 1 in the [A] color block [B] achromatic shape block and [C] colored shape block.

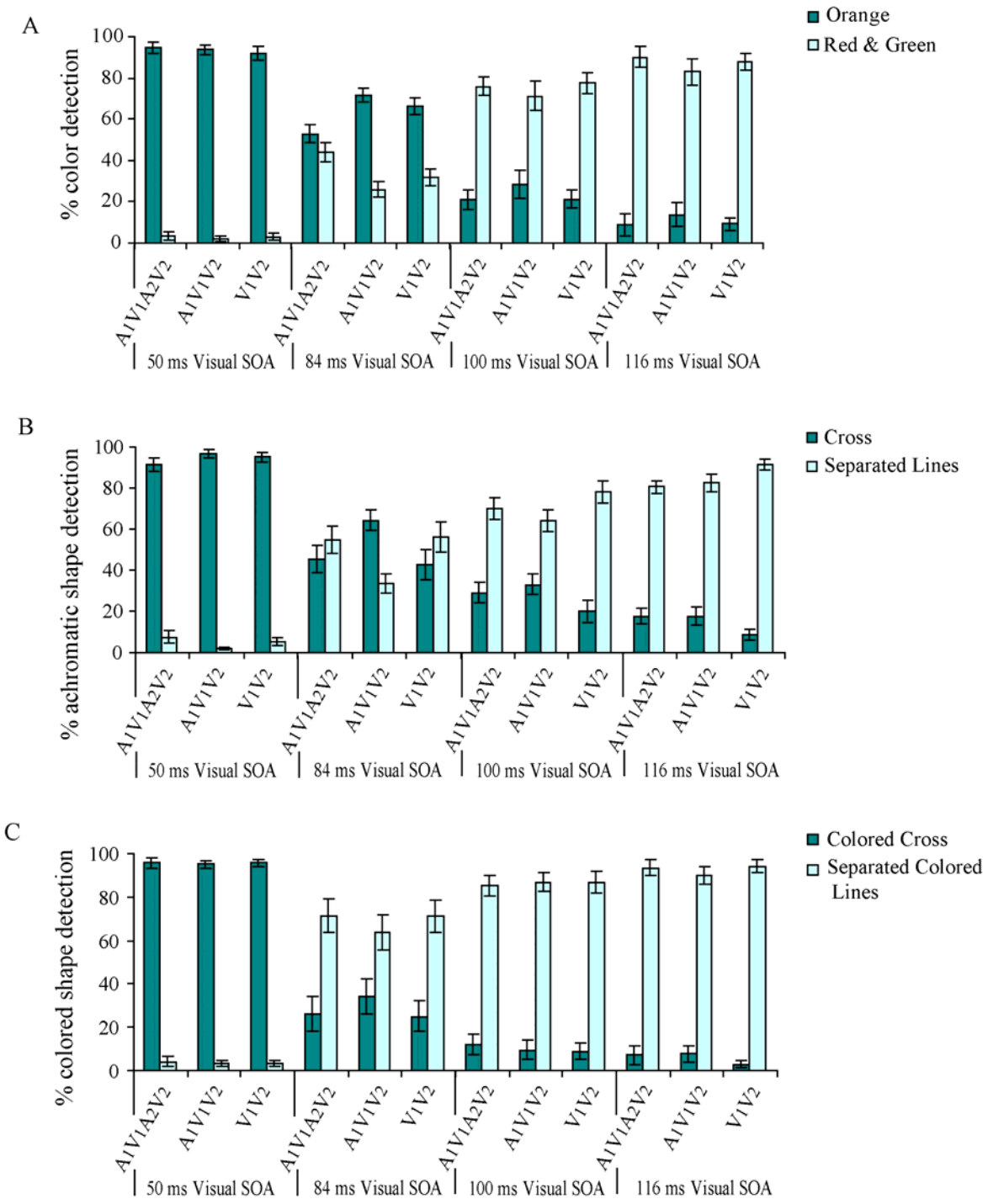

Figure 5.3 Fusion of sequential visual stimuli measured at increasing visual SOAs in the presence of (two or one) or absence of sounds for [A] colored stimuli (red and green merging to orange), [B] oriented lines (horizontal and vertical lines merging to a cross) and [C] oriented colored lines (merging to form a colored cross).

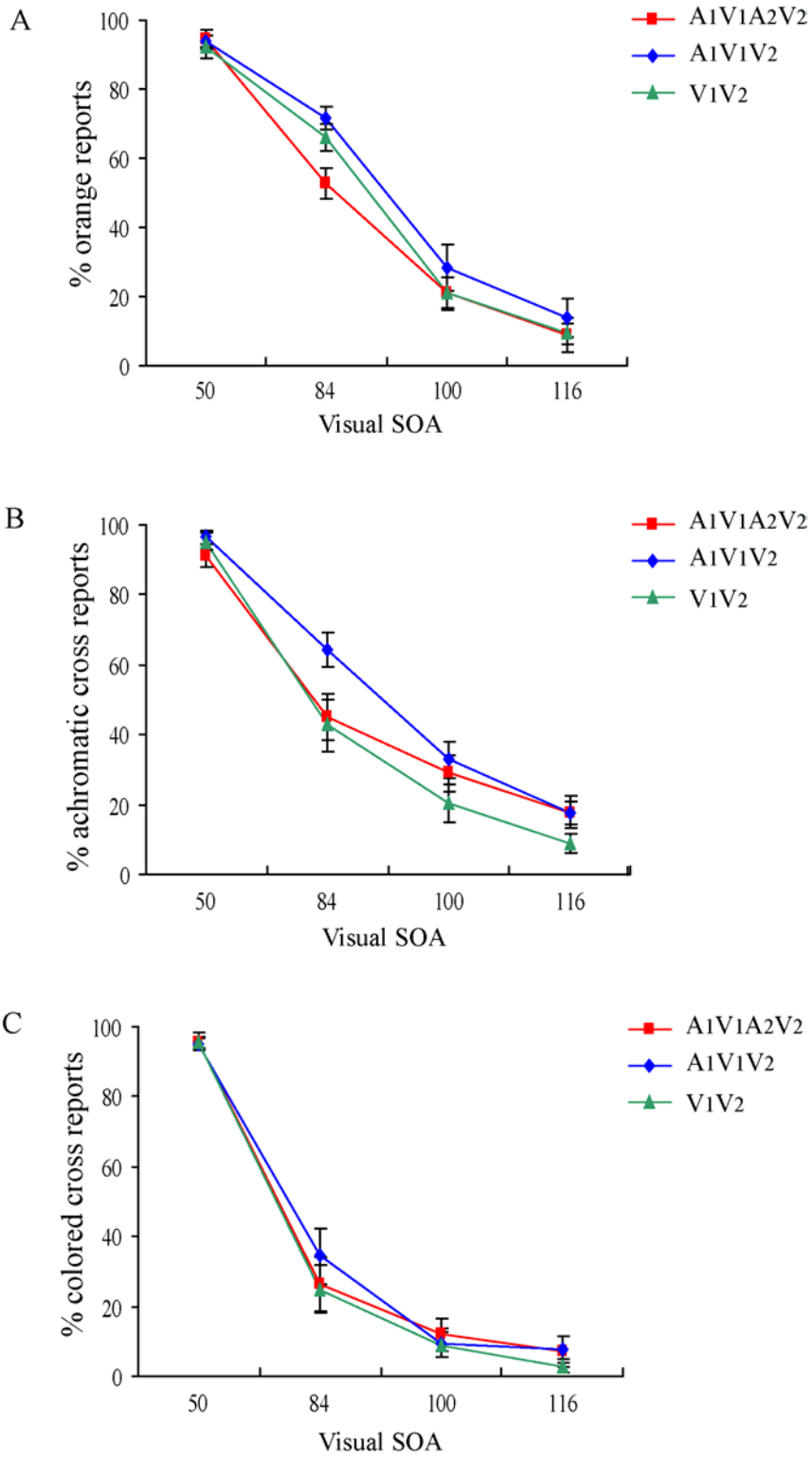

Figure 5.4 Fusion of sequential visual stimuli as a function of visual SOA for [A] colored stimuli, [B] achromatic lines and [C] colored lines.

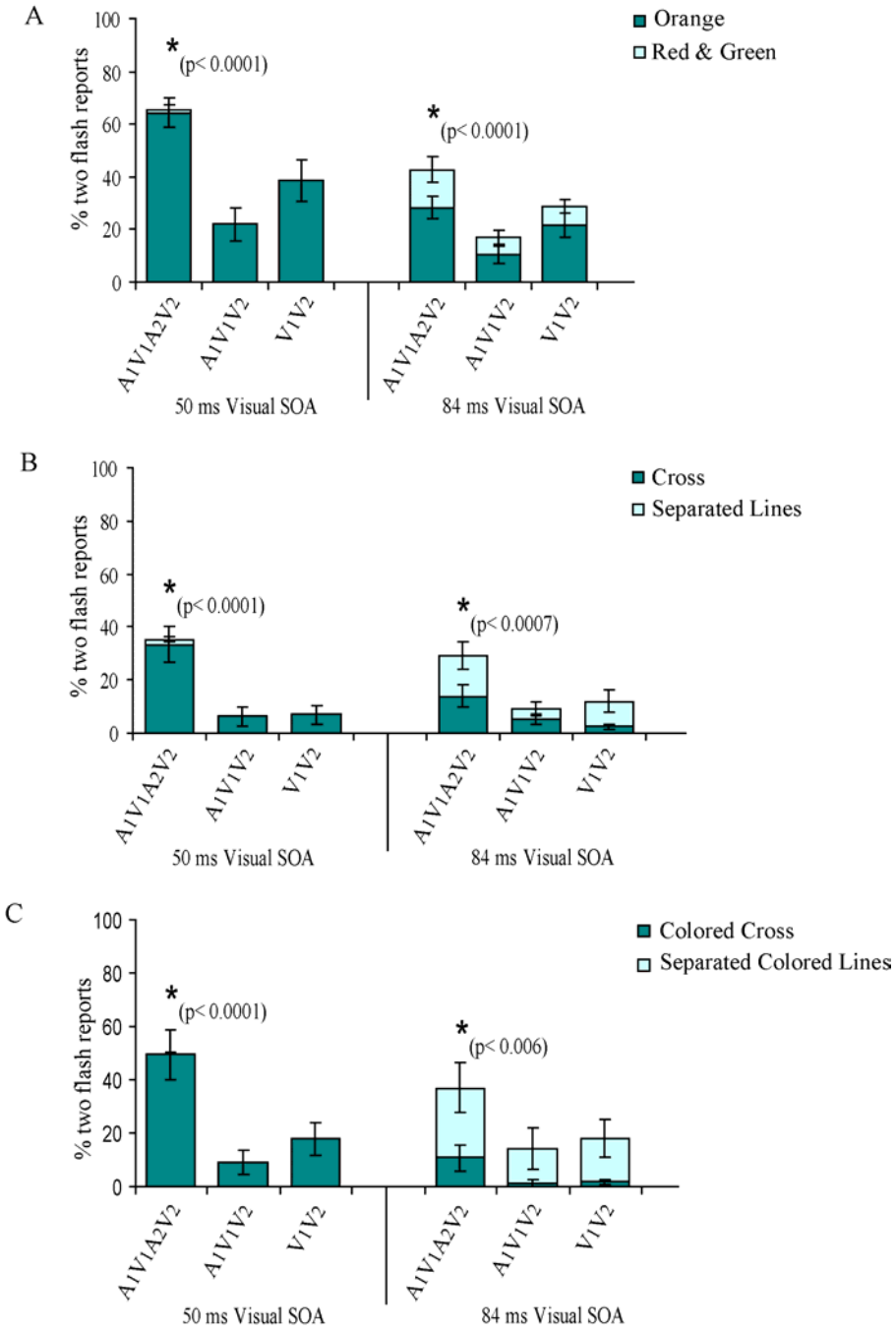

Figure 5.5 Percent two flash reports on audio-visual/ visual stimuli tested in Experiment 2 for the 50 ms and 84 ms V_1-V_2 SOAs in the [A] color block [B] achromatic shape block and [C] colored shape block. The 'orange', 'cross' and 'colored cross' in [A], [B] and [C] were not physically presented but perceived stimuli as a result of perceptual fusion of the first and second flash. Two flashes were reported on the separated colors in [A] or separate lines in [B] and [C] only when either one of the separated colors/ lines appeared to flash twice. '*' indicates significant differences in two flash reports on trials with two sounds vs. one or no sound at that SOA.

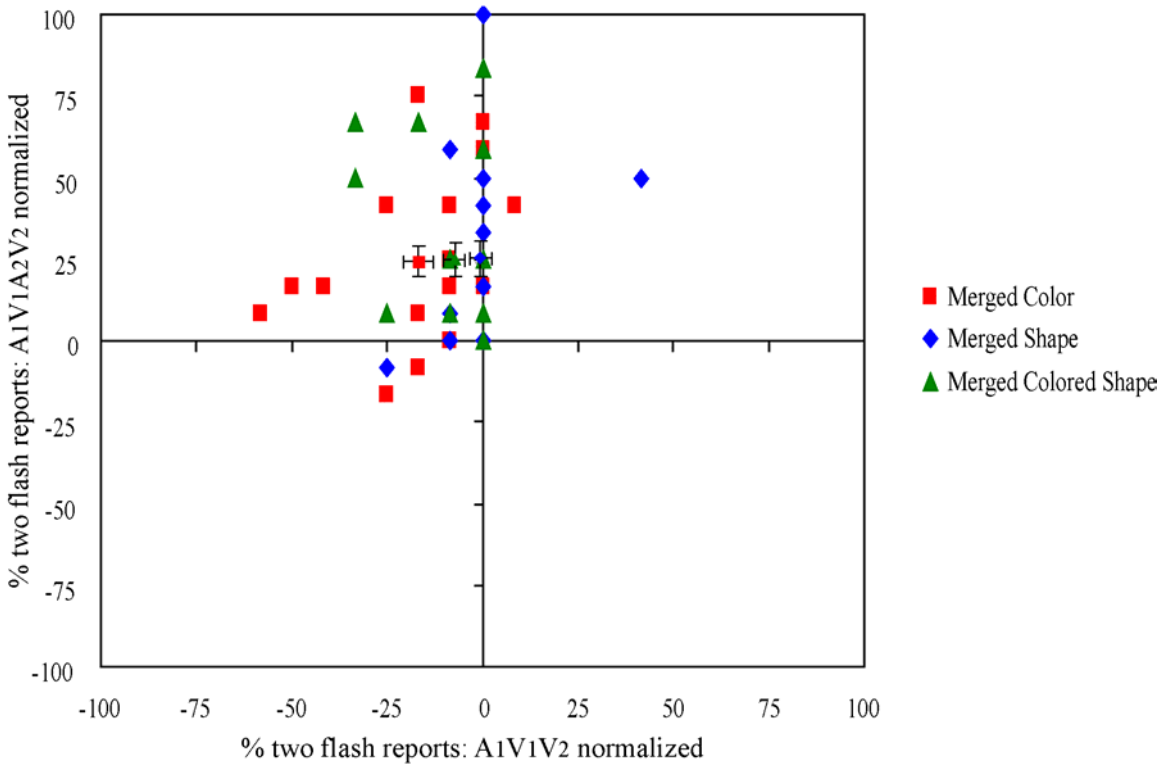

Figure 5.6 Two flash reports in Experiment 2 for the 50 ms V_1V_2 SOA on $A_1V_1A_2V_2$ trials in comparison to $A_1V_1V_2$ trials for the merged color, shape and colored shape stimuli. Each point represents data from an individual subject. Data points with error bars are the mean two flash reports and the standard error of the means for the three types of stimuli. Data points on both axes are normalized relative to two flash detection on the V_1V_2 stimulus configuration.

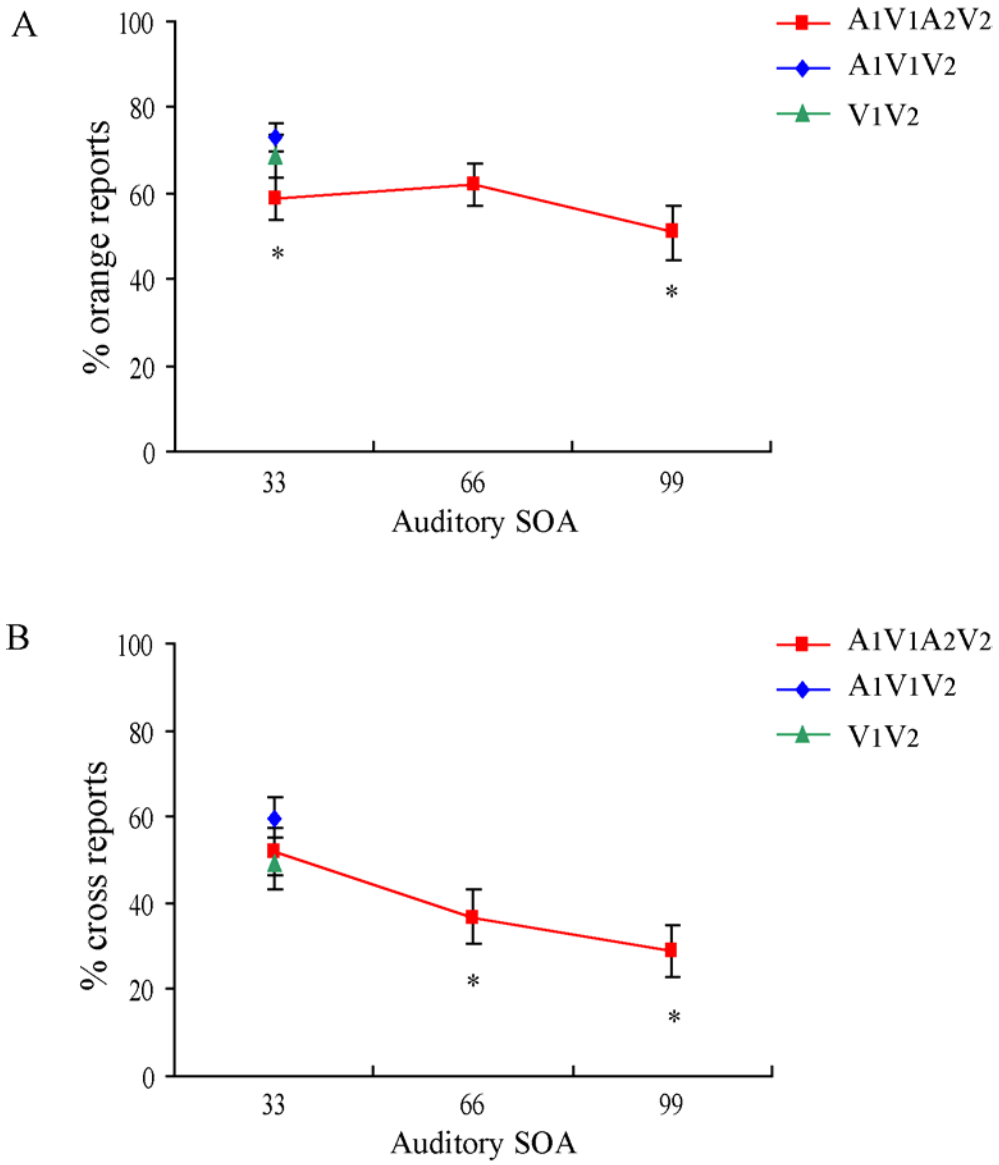

Figure 5.7 Fusion of sequential visual stimuli as a function of A_1A_2 auditory SOA for [A] colored stimuli and [B] achromatic lines. Auditory SOA were only valid for $A_1V_1A_1V_2$ stimuli and not for $A_1V_1V_2$ or V_1V_2 stimuli. '*' indicates significantly reduced fused percepts on trials with two sounds vs. one or no sound (statistics in text).

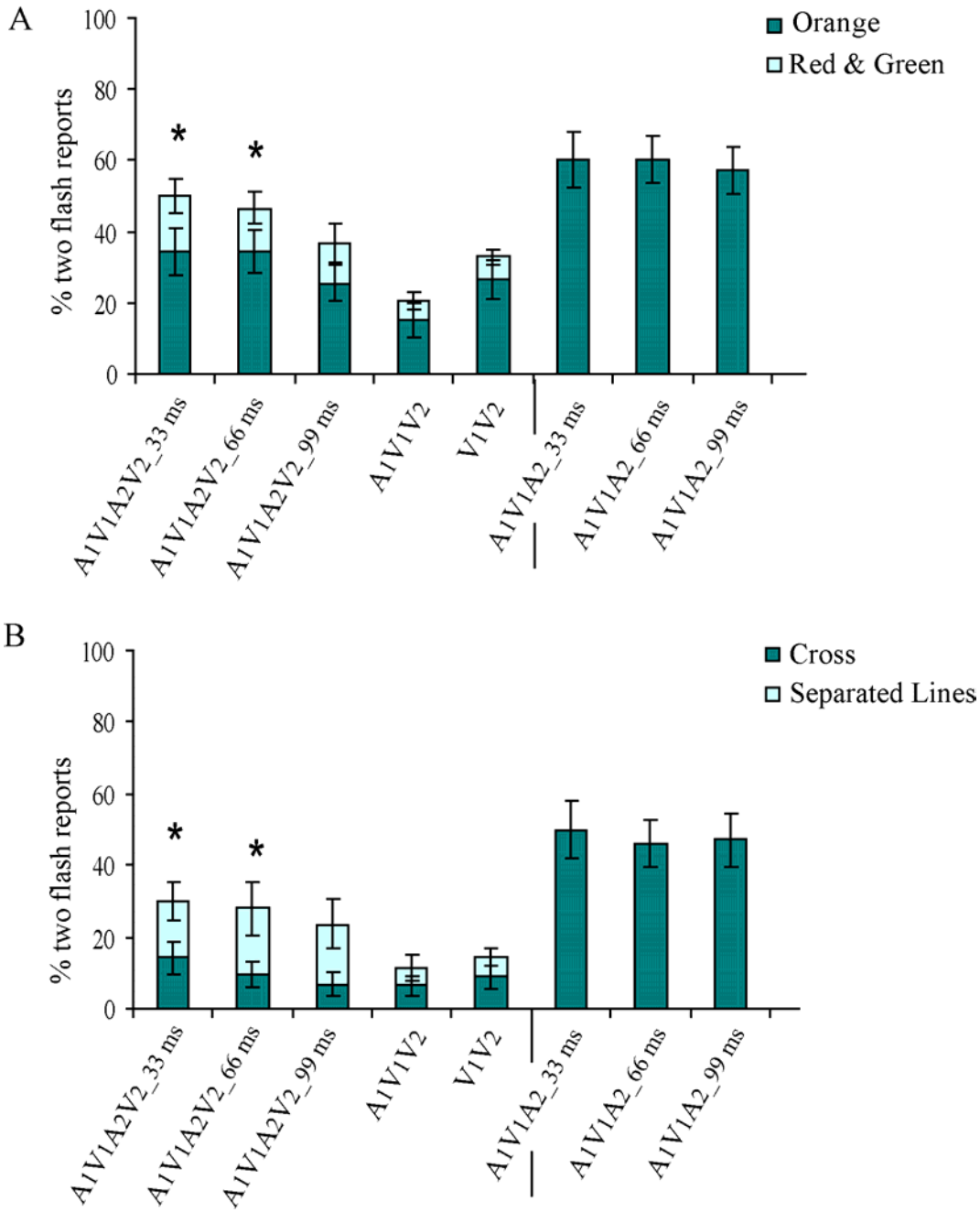

Figure 5.8 Percent two flash reports on audio-visual/ visual stimuli tested in Experiment 3 in the [A] color block and [B] achromatic shape block. '*' indicates significantly different two flash reports on trials with two sounds vs. one or no sound (statistics in text).

REFERENCES

Arden GB, Wolf JE, Messiter C (2003) Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. Vision Res 43:2469-2478.

Berger TD, Martelli M, Pelli DG (2003) Flicker flutter: is an illusory event as good as the real thing? J Vis 3:406-412.

Burr D, Alais D (2006) Combining visual and auditory information. Prog Brain Res 155:243-258.

King AJ, Calvert GA (2001) Multisensory integration: perceptual grouping by eye and ear. Curr Biol 11:R322-325.

Lippert M, Logothetis NK, Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. Brain Res 1173:102-109.

McCormick D, Mamassian P (2007) What does the illusory-flash look like? Vision Res.

Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci 27:4120-4131.

Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. Brain Res Cogn Brain Res 17:154-163.

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078-1093.

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

Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. Neuroreport 12:3849-3852.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147-152.

Shams L, Iwaki S, Chawla A, Bhattacharya J (2005) Early modulation of visual cortex by sound: an MEG study. Neurosci Lett 378:76-81.

Spence C, Squire S (2003) Multisensory integration: maintaining the perception of synchrony. Curr Biol 13:R519-521.

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge, MA: MIT.

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. Neuroimage.

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychol Bull 88:638-667.

Welch RB, DuttonHurt LD, Warren DH (1986) Contributions of audition and vision to temporal rate perception. Percept Psychophys 39:294-300.

Chapter 6: Neural Processes underlying an Auditory-induced Visual Illusion

DISCUSSION

The aim of the present thesis was to identify and characterize the neural basis of a cross-modal illusion termed the sound-induced extra flash illusion. The illusion is generated when a brief flash of light is interposed between two pulsed sounds. This audio-visual presentation produces the visual percept of two distinct flashes of which the second is illusory (Shams et al., 2001, 2002). Investigating the neural basis of this double-flash illusion provides a powerful approach for revealing how information from different modalities is integrated in the brain. As the illusion consists of a discrete visual perceptual event that varies on a trial-by-trial basis, it offers the possibility of isolating the critical sequence of neural events by which an auditory input induces a visual percept.

In this thesis, the underlying neural basis of the sound-induced illusory flash was investigated using 64 channel event related potential (ERP) recordings in conjunction with anatomical source localization. The visual properties of the illusory flash were further explored in a behavioral study. The main findings from these studies are summarized below:

Experiment I

Neural activity associated with the illusory flash was isolated in difference ERPs (Ill_Diff) calculated by subtracting the sum of the ERPs to the visual (V_1) and auditory (A_1A_2) stimuli presented alone from the ERP to the cross-modal combination ($A_1V_1A_2$). Ill_Diff = [($A_1V_1A_2$ + NoStim) – (A_1A_2 + V_1)].

In this experiment, the neural correlates of the cross-modal illusion were investigated with respect to

i) differences amongst individual subjects, those who perceived the illusion more frequently (SEE group) versus those who did not perceive the illusion often (NO-SEE group),

ii) differences on a trial-by-trial level within subjects, that is differences between trials on which the illusory flash was seen (SEE trials) versus not seen (NO-SEE trials).

The Ill_Diff waveforms revealed an early modulation of visual cortex activity at 30–60 ms after the second sound, which was larger in amplitude in subjects who saw the illusory flash more frequently. This modulation was termed the PD120, a positive deflection within the difference waveforms peaking at 120 ms post stimulus onset. Across subjects the PD120 amplitude showed a significant positive correlation with the proportion of trials on which the illusion was seen.

The trial-by-trial analysis of the ERPs did not find any difference in the occipital PD120 component. Instead the analysis revealed an enlarged negativity in the 90-150 ms range in the crossmodal interaction waves corresponding to the SEE versus NOSEE trial difference. This trial specific component localized to auditory cortex in its early phase (ND110) and to polymodal superior temporal cortex in its late phase (ND130). Concurrent gamma bursts (25-40 Hz) during the 110-145 ms (EP130) and 200-240 ms (EP220) latencies were found in visual cortex on SEE trials associated with perception of the double-flash illusion.

These findings suggested that the perception of the illusory second flash is based on a rapid dynamic interplay between modality-specific auditory and visual cortical areas

as well as polymodal superior temporal cortex. This neural activity pattern associated with the auditory induced illusory flash differed markedly from that evoked by a real second flash. The difference wave components to a real second flash (isolated as $V_2 = (V_1\text{-}V_2) - V_1$) consisted of the standard sequence of C1-P1-N1 components that is well-documented to be evoked by visual stimuli (Di Russo et al., 2002, 2003; Clark and Hillyard, 1996). Thus, contrary to earlier proposals (Shams et al., 2001), there appears to be little overlap between the neural events that underlie real and illusory flashes.

Experiment II

This study investigated the neural basis of the sound-induced flash fusion phenomenon, which is the complement of the sound induced extra flash illusion (Andersen et al., 2004, Shams et al., 2005, Watkins et al., 2007). This cross-modal phenomenon is a perceptual fusion of two sequential flashes when presented concurrent with one sound, so that only one flash is reportedly seen. This behavioral effect was demonstrated as part of Experiment I. Additional analyses of the data obtained in Experiment I were undertaken to identify the neural processes underlying flash fusion. The neural correlates of the effect emerged 80-110 ms after the second flash in the form of a positive difference wave component, PD180, which was markedly attenuated in subjects who did not perceive the second flash. This PD180 was localized to polysensory superior temporal cortex. The trial based analysis found that the PD180 along with a subsequent modulation in visual cortex at 228-248 ms were diminished on fusion trials relative to trials when two flashes were correctly reported. These results suggested that

sound induced flash fusion was based on an interaction between polysensory and visual cortical areas.

The trial based temporal dynamics underlying flash fusion, however, differed from the neural correlates of the extra flash illusion. The PD180 component was found to be associated with both phenomena, but it varied as a function of perceptual reports only in the case of flash fusion. The PD180 within the fusion difference waveforms was also found to covary in amplitude across subjects with the visual evoked N1 component (148-184 ms), whose early phase (148-168 ms) localized to the same cortical region as the PD180. Together, these results suggest that the activation in the superior temporal region giving rise to the PD180 is related more to the precise timing and segmenting of visual inputs than to the generation of an illusory visual percept.

Experiment III

This study investigated the effect of spatial attention on the sound-induced extra flash illusion. Attention was focused upon stimuli at either one of two locations in space in the upper (UVF) or lower visual field (LVF). The extra flash illusion was reproducible in response to $A_1V_1A_2$ stimuli at either location. Attention was found to enhance all difference wave components in the Ill_Diff waveform (calculated as in Experiment I). Importantly, attention was crucial to the generation of the early PD120 component, as the component was insignificant in the unattended Ill_Diff waveforms. This suggested that attention maybe a pre-requisite for the generation of the illusion and that the perception of the illusion is not an automatic multisensory integration process.

Within this experiment, PD120 was observed to have a positive polarity for stimuli in either the UVF or LVF and was localized to the ventral occipito-temporal extrastriate visual cortex, similar to its localization in Experiment I. This visual area has been shown to be modulated by attention in numerous ERP and fMRI studies (Kastner et al., 1998, Corbetta et al., 2000, Hopfinger et al., 2000, Martinez et al., 2001, 2006, 2007). The localization of PD120 to this region and its susceptibility to attention suggest that the neural populations that are modulated by top-down attention in these regions may also be the recipients of auditory inputs that can alter visual sensation.

Of note, along with the extrastriate source localization of PD120, the finding that the component does not invert in polarity for stimuli in the upper vs. lower field further suggests that primary visual (striate) cortex activity is not a strong contributor to its generation (Clark and Hillyard, 1996, Di Russo et al., 2002, 2003).

Experiment IV

This behavioral study investigated the sensory properties of the illusory extra flash elicited when the real flash had simple attributes such as color and basic line shapes. In all cases, visual stimuli with these features were susceptible to the extra flash illusion suggesting that the illusion is not specific to a particular type of visual 'object'.

The sensory properties of the illusory flash were probed in the context of perceptual fusion of sequentially flashed stimuli, wherein the first and second visual stimuli had different feature attributes (such as different colors or different line orientations) that fused to generate a new percept having a blend of the attributes (such as a new color or shape). Our original hypothesis was that the perceptual fusion between the

features of the first and second flashed stimulus may be qualitatively enhanced due to the co-occurrence of the extra flash illusion to the first real flash at the time of the second flash. On the contrary, it was found that perceptual fusion between sequential visual stimuli was reduced on two-sound trials relative to the presence of one or no sounds, which is a novel finding.

It was also found that the illusory flash was consistently generated to the fused percept resulting from the merging of sequential visual stimuli. As perceptual fusion of successively presented features decreased with increasing visual stimulus separations, the frequency of the extra flash illusion also decreased. That the extra flash illusion is generated following the fusion of sequential stimuli suggests that the auditory induced effect may be manifested at higher levels in the visual processing hierarchy beyond those at which fusion of the colored/ oriented line stimuli occurs. This reasoning would be consistent with the visual physiology of the illusion studied in the previous experiments, which suggested that the crossmodal interaction underlying the illusion occurs outside of primary visual (striate) cortex.

A NEURAL MODEL

From the cumulative results of the above experiments, a neural model can be hypothesized for the sound-induced extra flash illusion:

i) Voluntary attention to the stimulus location primes the sensory cortical regions to incoming audio-visual inputs (Womelsdorf and Fries, 2006, Reynolds and Chelazzi, 2004, Treue, 2001).

ii) On trials when the illusion is perceived, the second sound (A_2) sets off a rapid cross-modal interaction in the auditory cortex (ND110) occurring within 20-40 ms after A_2. The effect of A_2 is to enhance the stimulus processing of the first sound (A_1) within auditory cortex in the auditory N1 latency range. This provides the earliest trigger leading to perception of the illusion.

iii) In individuals who are disposed to see the illusion the $A_1V_1A_2$ cross-modal interaction produces an early response in visual cortex, PD120, onsetting 30-60 ms after A_2. PD120 is predominantly generated within extrastriate cortex in the ventral visual stream and represents an enhanced neural response that depends on attention. This neural modulation, however, is necessary but not sufficient for seeing the illusory flash.

iv) On trials when the illusion is perceived, the early cross-modal interactions in modality specific cortices, ND110 and PD120, act synergistically, producing an enhanced burst of gamma band EEG power in visual cortex (EP130) immediately following these interactions.

v) The above cortical interplay in unisensory areas is almost simultaneously relayed to polymodal cortex in the superior temporal area, where an increased negativity is observed on SEE trials, ND130.

The rapid interaction between the unisensory auditory and visual cortices described here may be engendered by direct connections between these areas. Recent anatomical labeling studies in macaques (Falchier et al., 2002; Clavagnier et al., 2004; Rockland and Ojima, 2003) have identified projections from primary auditory (AI) and auditory association cortices to the visual cortex (areas V1 and V2), as well as from

visual to auditory cortices (Bizley et al., 2007). In all cases, projections into primary sensory areas, V1 and A1, were found to be sparse compared to denser innervations into higher areas in the sensory processing hierarchy. These findings are in agreement with the source localization of the PD120 component which was found to predominantly arise from activity within extrastriate visual areas. It has been hypothesized that the function of these early connections may involve modulations related to weighting of one modality relative to another, or alerting, but not specific relations such as semantic/ associative links between modalities that may require longer/ more complex processing (Driver and Noesselt, 2008).

The experiments presented in this thesis suggest that the illusion is associated with a perturbation of activity within visual cortex. The illusory flash, however, is generated in a unique audio-visual context by a conjunction of activations in different cortical regions. The neurophysiological basis as well as the psychophysical properties of the illusory flash were thus found to differ markedly from those of a real flash.

FUTURE DIRECTIONS

Further questions can be explored pertaining to the illusory flash phenomenon to pin down its specific neural mechanisms. For instance, the second auditory stimulus was found to be the critical trigger for the generation of the illusion. The effect of changing the properties of this stimulus, such as a different frequency or intensity from that of the first auditory stimulus, on the illusion and its correlates can be explored. Spatial congruity of the auditory and visual components of the $A_1V_1A_2$ stimulus is an important aspect of the illusion. That is, the illusion has been reported to break down when the

A_1A_2 and V_1 components are spatially separated (Shams et al., 2002). The effects of spatial incongruity on the early cortical dynamics underlying the processing of $A_1V_1A_2$ can be delineated.

Broader questions related to the phenomenon include investigation of parallel effects in visual-somatosensory conjunctions. Violentyev et al. (2005) showed that two finger taps in conjunction with a flash can also generate the percept of two flashes, although less effectively than the visual illusion induced by two sounds. Investigation of the neural basis of the somatosensory-visual illusion can help delineate common mechanisms across somatosensory-visual and audio-visual interactions.

Another intriguing question is how plastic or modifiable are the audio-visual neural connections that mediate the illusion; that is can they be perturbed via short-term or long-term audio-visual associations? For instance, exposure to associations between temporally disparate auditory and visual stimuli can recalibrate audio-visual simultaneity judgments (Fujisaki et al., 2002). Similarly, if subjects are constantly exposed to either A_1V_1 or $A_1V_1A_2V_2$ stimulus associations in an experimental session, can that alter the illusory percept and its underlying neural correlates on subsequent $A_1V_1A_2$ trials? With respect to long-term audio-visual exposure, it is an interesting question whether individuals who have had long-term audio-visual training, such as musicians, have similar neural mechanisms underlying the illusion as found here. The present studies found that amongst normal individuals there is a huge variability in the frequency with which the illusion is seen. These behavioral differences were neurally manifest in the PD120 component, which was found to differ between the SEE and NO-SEE subject groups not only for the $A_1V_1A_2$ stimulus but also for audio-visual stimuli such as

$A_1V_1A_2V_2$ that do not produce an illusion. These findings suggest consistent individual differences in the neural connectivity that underlies cross-modal integration. Studying the anatomy and physiology of intersensory neural connectivity that underlies individual differences on cross-modal perceptual phenomena can shed light on the neural bases of crossmodal integration in normal individuals and even in synesthetic individuals who make abnormal cross-modal associations (Beeli et al., 2008, Hubbard, 2007).

REFERENCES

Andersen TS, Tiippana K, Sams M (2004) Factors influencing audiovisual fission and fusion illusions. Brain Res Cogn Brain Res 21:301-308.

Beeli G, Esslen M, Jancke L (2008) Time course of neural activity correlated with colored-hearing synesthesia. Cereb Cortex 18:379-385.

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17:2172-2189.

Clark VP, Hillyard SA (1996) Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. J Cogn Neurosci 8:387-402.

Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. Cogn Affect Behav Neurosci 4:117-126.

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci 3:292-297.

Di Russo F, Martinez A, Sereno MI, Pitzalis S, Hillyard SA (2002) Cortical sources of the early components of the visual evoked potential. Hum Brain Mapp 15:95-111.

Di Russo F, Martinez A, Hillyard SA (2003) Source analysis of event-related cortical activity during visuo-spatial attention. Cereb Cortex 13:486-499.

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57:11-23.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749-5759.

Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. Nat Neurosci 7:773-778.

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3:284-291.

Hubbard EM (2007) Neurophysiology of synesthesia. Curr Psychiatry Rep 9:193-199.

Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282:108-111.

Martinez A, DiRusso F, Anllo-Vento L, Sereno MI, Buxton RB, Hillyard SA (2001) Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. Vision Res 41:1437-1457.

Martinez A, Teder-Salejarvi W, Vazquez M, Molholm S, Foxe JJ, Javitt DC, Di Russo F, Worden MS, Hillyard SA (2006) Objects are highlighted by spatial attention. J Cogn Neurosci 18:298-310.

Martinez A, Teder-Salejarvi W, Hillyard SA (2007) Spatial attention facilitates selection of illusory objects: evidence from event-related brain potentials. Brain Res 1139:143-152.

Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. Annu Rev Neurosci 27:611-647.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19-26.

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

Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. Neuroreport 12:3849-3852.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147-152.

Shams L, Ma WJ, Beierholm U (2005) Sound-induced flash illusion as an optimal percept. Neuroreport 16:1923-1927.

Treue S (2001) Neural correlates of attention in primate visual cortex. Trends Neurosci 24:295-300.

Violentyev A, Shimojo S, Shams L (2005) Touch-induced visual illusion. Neuroreport 16:1107-1110.

Watkins S, Shams L, Josephs O, Rees G (2007) Activity in human V1 follows multisensory perception. Neuroimage 37:572-578.

Womelsdorf T, Fries P (2006) Neuronal coherence during selective attentional processing and sensory-motor integration. J Physiol Paris 100:182-193.