BRAIN RESEARCH

Research Report

Cortical processes underlying sound-induced flash fusion

Jyoti Mishra[a,*], Antigona Martinez[a,b], Steven A. Hillyard[a]

[a]Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA
[b]Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

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, T.S., Tiippana, K., Sams, M., 2004. Factors influencing audiovisual fission and fusion illusions. Brain Res. Cogn. Brain Res. 21, 301–308; 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., Ma, W.J., Beierholm, U., 2005b. Sound-induced flash illusion as an optimal percept. Neuroreport 16, 1923–1927]. 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, J., Martinez, A., Sejnowski, T.J. and Hillyard, S.A., Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J. Neurosci. 27 (2007) 4120–4131]. 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 160–190 ms after stimulus onset, which was markedly attenuated in subjects who did not perceive the second flash. This component, previously designated as PD180 [Mishra, J., Martinez, A., Sejnowski, T.J. and Hillyard, S.A., Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J. Neurosci. 27 (2007) 4120–4131], 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 inter-individual differences in perceiving the illusion are based at least in part 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 two flashes were correctly reported. These results suggest that the sound induced flash fusion is based on an interaction between polysensory and visual cortical areas.

Published by Elsevier B.V.

1. 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 multi-modal 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

Table 1 – Mean behavioral performance for reporting the number of flashes seen (one or two) for stimulus combinations containing one or two visual stimuli				
Stimulus	Percent correct discrimination of number of flashes	SEM (% trials)	Mean RT (ms)	SEM RT (ms)
V_1	87	1.9	612	11
V_1V_2	67	3.5	660	13
A_1V_1	91	1.1	591	14
$A_1V_1A_2V_2$	87	1.7	615	14
$A_1V_1V_2$	56	5.2	663	12
$A_1V_1A_2$	63	4.2	684	12
$A_1A_2V_1$	91	1.1	581	15

Percent trials on which the number of stimulus flashes were correctly reported and the standard error of these percentages (SEM) are reported over all 34 subjects. Mean response time (RT) measures and the standard error of these RTs over all subjects are also shown (data from Mishra et al., 2007).

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 de 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. This phenomenon has been observed in some previous behavioral studies (Andersen et al., 2004; Shams et al., 2005b), but was absent in others (Shams et al., 2002; Meylan and Murray, 2007). Recently, the flash fusion effect was studied in an fMRI investigation, which showed that modulation of primary visual cortex may accompany 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. The study was performed in a large cohort of subjects,

which made it possible to investigate the underlying neural mechanisms in individuals who perceived the flash fusion effect versus those who did not. Accordingly, we studied the spatio-temporal patterns of neural activity associated with the flash fusion percept by making both between-subject comparisons and within-subject comparisons on a trial-by-trial basis. 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. The data in the present study were obtained as part of a broader ERP study that investigated not only the flash fusion effect but also the extra flash illusion generated by a two-sound-one-flash stimulus as well as other non-illusory cross-modal interactions within the same design (Mishra et al., 2007). The analyses of these data thus allowed comparisons of the neural correlates of different types of illusory and non-illusory intersensory interactions.

2. Results

2.1. Behavioral results

The experimental layout and the different auditory (A), visual (V) and audio-visual (AV) stimulus configurations presented to subjects in visual periphery are shown in Fig. 7. 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 are shown in 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 (SEM 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
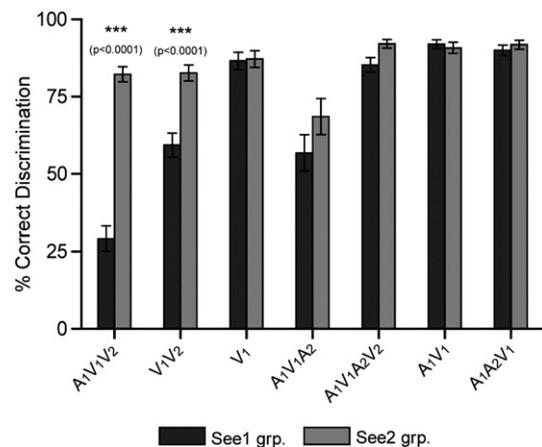

Fig. 1 – Behavioral performance comparisons across all experimental stimuli between subjects who frequently perceived the two flash component of the $A_1V_1V_2$ stimulus as a single flash (SEE1 group), and those who correctly reported seeing two flashes on the majority of trials (SEE2 group).
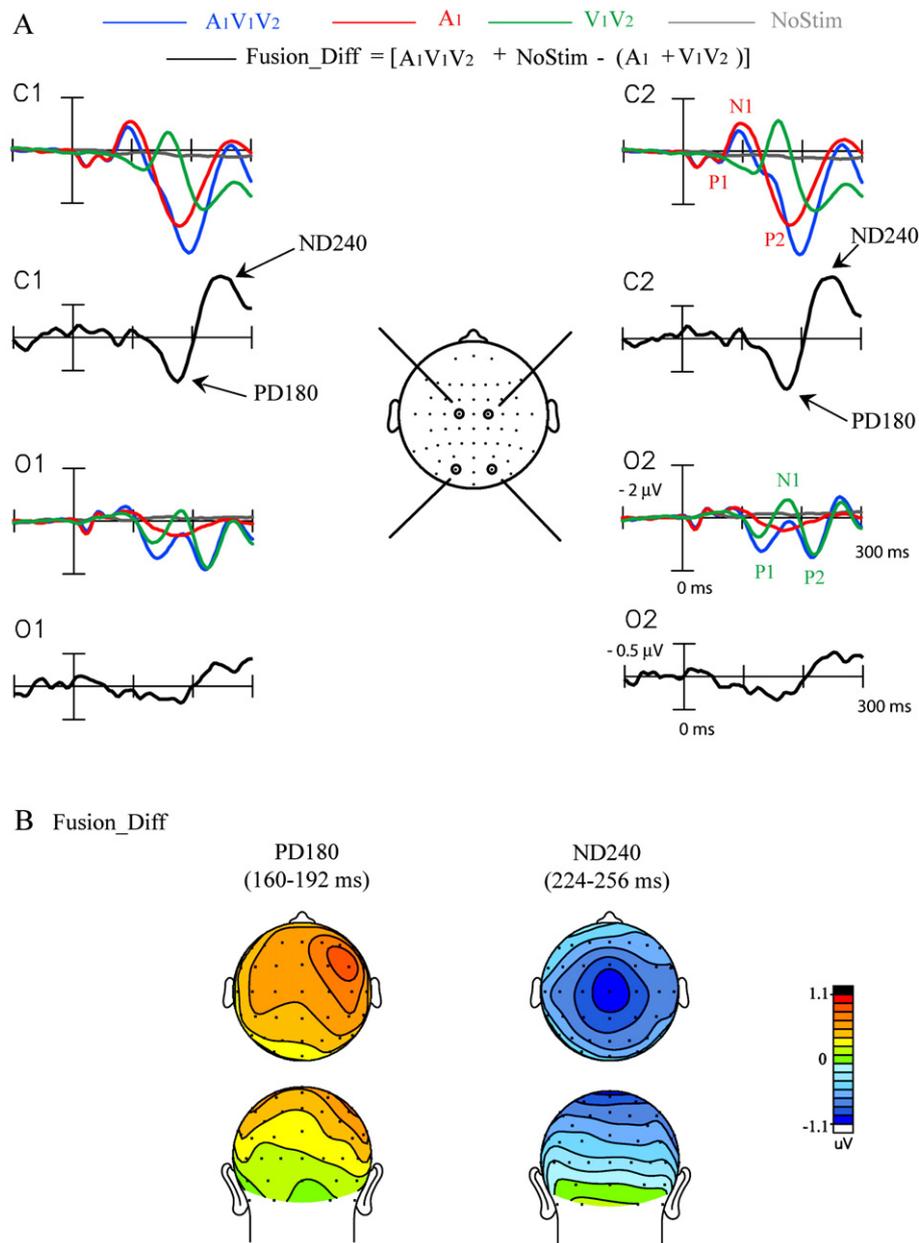
Fig. 2 – Grand-average ERPs ($n=34$) associated with the sound-induced flash fusion illusion. [A] ERPs elicited by the illusion-inducing $A_1V_1V_2$ stimulus and by its unimodal constituents A_1 and V_1V_2, together with the ERP time-locked to the blank "No-Stim" event. The Fusion_Diff difference waves 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.

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%. 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.

Fig. 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.). The experimental design also included $A_1A_2V_1$ catch trials that were stimulus matched to the $A_1V_1A_2$

illusory stimulus. Within $A_1A_2V_1$ the visual flash (V_1) was dissociated from the auditory A_1A_2 component by a 200 ms delay, rendering the stimulus non-illusory. Individuals in both SEE1 and SEE2 groups correctly discriminated the $A_1A_2V_1$ stimulus as containing a single flash (Fig. 1, 90% SEE1 grp. vs. 92% SEE2 grp., $t(32)=0.81$, $p=$n.s.). These results suggest that group differences in SEE1/SEE2 responses were based on actual perceptual experience rather than a response bias to report the number of flashes based on the number of sounds.

To further demonstrate that the behavioral differences between the SEE1 and SEE2 groups in their responses to the $A_1V_1V_2$ and V_1V_2 stimuli were due to differences in perceptual sensitivity rather than response or decision bias, signal detection estimates of sensitivity (d') and decision criterion (ß) were calculated (see Methods). Perceptual sensitivity in the SEE1 group was significantly lower than in the SEE2 group for both the $A_1V_1V_2$ (SEE1 vs. SEE2 d': 0.91 vs. 2.53, $t(32)=6.89$, $p<0.0001$) and V_1V_2 (SEE1 vs. SEE2 d': 1.49 vs. 2.30, $t(32)=3.42$, $p<0.002$) stimuli. Decision criteria, however, did not differ between the two groups for either the $A_1V_1V_2$ (SEE1 vs. SEE2 ß: 3.28 vs. 3.34, $t(32)=0.05$, $p=$n.s.) or V_1V_2 (SEE1 vs. SEE2 ß: 0.94 vs. 0.68, $t(32)=0.85$, $p=$n.s.) stimuli.

Across all subjects a significant correlation was found 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 had a general propensity to perceive rapid double flashes as unitary. Importantly, this propensity was not completely responsible for the flash fusion perception of the $A_1V_1V_2$ stimulus, since the presence of the A_1 sound significantly increased the perceptual reports of fusion (SEE1 group: 41% flash fusion on V_1V_2 and 71% fusion on $A_1V_1V_2$; stimulus condition × group interaction: $F(1, 32)=38.52$, $p<0.0001$).

2.2. 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 significant positive (P) and negative (N) deflections in these difference waves will be referred to as "components" for simplicity. 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 2. Components occurring after 300 ms in the Fusion_Diff waves were not analyzed because of the likelihood

Table 2 – Mean amplitudes of ERP components in the difference waves associated with sound-induced flash fusion (Fusion_Diff) averaged over all 34 subjects

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

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.

that activity related to decision making and response preparation would be confounded with activity related to cross-modal interaction and perceptual processing (Hillyard and Picton, 1987, Coles et al., 1995).

2.3. 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). In the Fusion_Diff waveforms, the PD180 component was found to be significantly larger in amplitude in the SEE2 vs. the SEE1 group ($F(1,32)=7.21$, $p<0.02$) (Figs. 3B and C). For the SEE1 group the PD180 mean amplitude did not even reach statistical significance with respect to pre-stimulus baseline (Table 3). 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 × Electrode interaction: $F(37, 1184)=1.49$, $p<0.04$), but this difference most likely arose because PD180 amplitude was near noise levels in the SEE1 group. No group differences were found in the topography of the ND240 component ($F(37,1184)=0.25$, $p=$n.s.). Of note, the two groups did not differ in their electro-ocular responses to the $A_1V_1V_2$ stimulus (Fig. 3A, HEOG and VEOG) indicating that sound-evoked blinks were not responsible for the behavioral differences between the groups.

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.
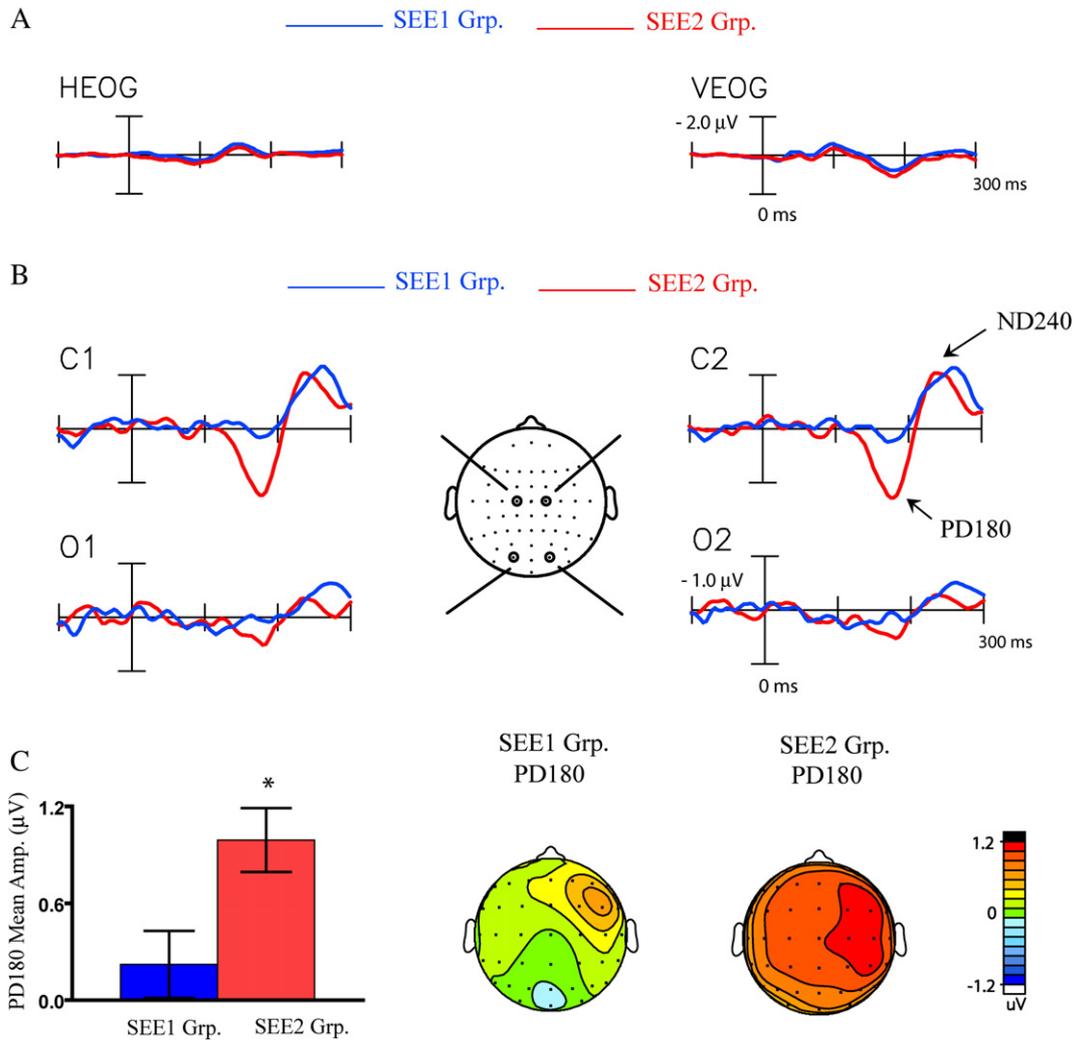
Fig. 3 – ERP differences between the SEE1 and SEE2 groups. [A] Horizontal and vertical electro-oculograms (HEOG and VEOG) time-locked to the $A_1V_1V_2$ stimulus and averaged separately for the SEE1 group (n = 17) and the SEE2 group (n = 17) [B] Fusion_Diff difference waves for the SEE1 and SEE2 groups. 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.

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. 4 and Table 3). The early phase of the visual evoked N1 (latency range 148–168

Table 3 – 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

ERP	Component	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
$A_1V_1V_2$	PD180 interval	2.69	0.42	6.46	0.0001	2.49	0.39	6.30	0.0001
V_1V_2	N1early (148–168 ms)	−0.75	0.21	3.65	0.003	−1.54	0.26	5.91	0.0001
	N1late (168–188 ms)	−0.72	0.30	2.42	0.03	−1.34	0.32	4.16	0.0008
V_1	N1early (148–168 ms)	−0.79	0.22	3.59	0.003	−1.49	0.25	5.98	0.0001
	N1late (168–188 ms)	−0.82	0.26	3.14	0.007	−1.47	0.31	4.75	0.0003

Components were measured over scalp sites of maximal amplitude and tested for significance with respect to the 100 ms pre-stimulus baseline.
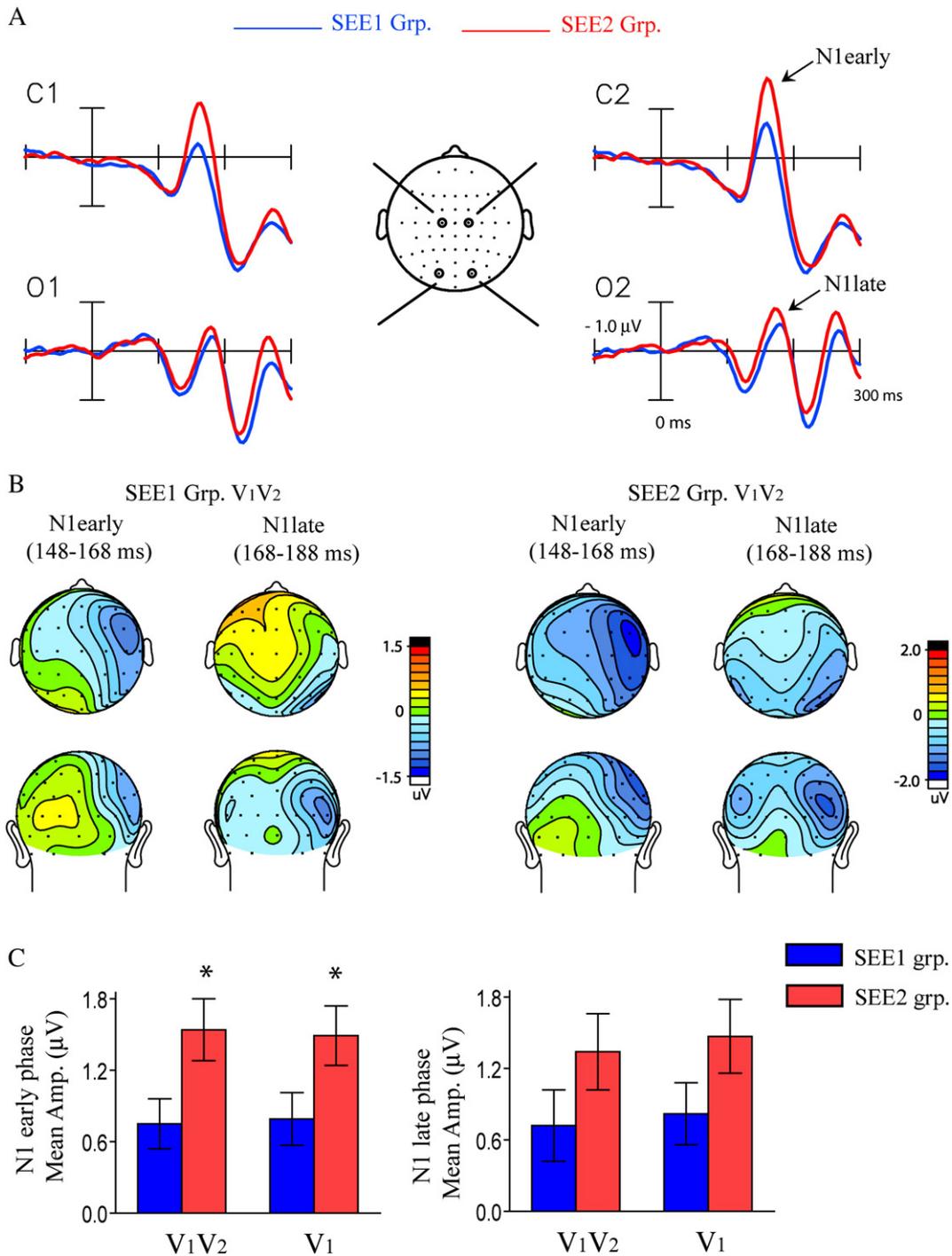
Fig. 4 – ERP differences between the SEE1 and SEE2 groups for the V_1V_2 stimulus. [A] ERPs to V_1V_2 averaged separately for the SEE1 group and the SEE2 group. Recordings are from left and right central (C1, 2) and occipital (O1, 2) sites. [B] Voltage maps comparing the topography of the visual N1 component in its early and late phases between the two groups. [C] Bar graphs comparing the mean amplitude of the early and late phases of the N1 component between the two groups in the ERPs to both the V_1V_2 and V_1 stimuli. '*' denotes significant amplitude differences between groups as reported in the text.

ms), which had a voltage maximum over anterior sites, was found to be significantly smaller for the SEE1 group compared to the SEE2 group ($F(1,32)=5.64$, $p<0.03$). A similar group difference was also found for the early phase of the N1 evoked by the single flash (V_1) stimulus ($F(1,32)=4.42$, $p<0.05$) (Fig. 4C, Table 3). The SEE1 vs. SEE2 group difference approached but

did not reach significance for the late phase of the N1 (168–188 ms), which had a contralateral occipital maximum, either for the double flash (V_1V_2) ($F(1,32)=2.80$, $p=$n.s.) or the single flash (V_1) stimulus ($F(1,32)=3.19$, $p=$n.s., Fig. 4C). The scalp topographies of the N1 component in the ERPs to V_1 vs. to V_1V_2 did not differ over the entire N1 interval (148–188 ms)

Table 4 – 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

	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)
	ND240	±45	−29	−6	Vicinity of STG	4% (224–256 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	$PD180_{Diff}$	±46	−6	−5	Vicinity of STG	10% (172–200 ms)
difference	$ND240_{OccDiff}$	±29	−64	−1	Vicinity of lingual gyrus	4% (228–248 ms)

Percent residual variance not accounted for by the model over the interval specified in parentheses is shown for each component.

(Condition×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×Electrode interaction: V_1V_2: $F(37,1184)=0.36$, $p=$n.s.; V_1: $F(37,1184)=0.15$, $p=$n.s.). These highly lateralized topographic distributions that did not differ between the two subject groups provided further evidence that the subjects in the two groups maintained central fixation to the same extent and hence viewed the stimuli at the same location in their visual periphery.

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 early N1 to the V_1V_2 stimulus and the PD180 amplitude in the Fusion_Diff waveform ($r(32)=-0.67$, $p<0.0001$). A relatively weaker correlation was also observed for the late phase of the N1 to V_1V_2 and the PD180 component ($r(32)=-0.45$, $p<0.008$). As shown in Table 3, the magnitude of the SEE2–SEE1 group difference for the PD180 component was 0.77 μV. This was comparable to the magnitude of the V_1V_2 evoked N1 group difference measured within the same latency range (160–192 ms) and over the same electrode sites as the PD180 component (0.74 μV). Thus, the differences in amplitude of the visual evoked N1 between the SEE1 and SEE2 groups might have contributed substantially to the group difference observed for the cross-modal PD180 component. Within this latency
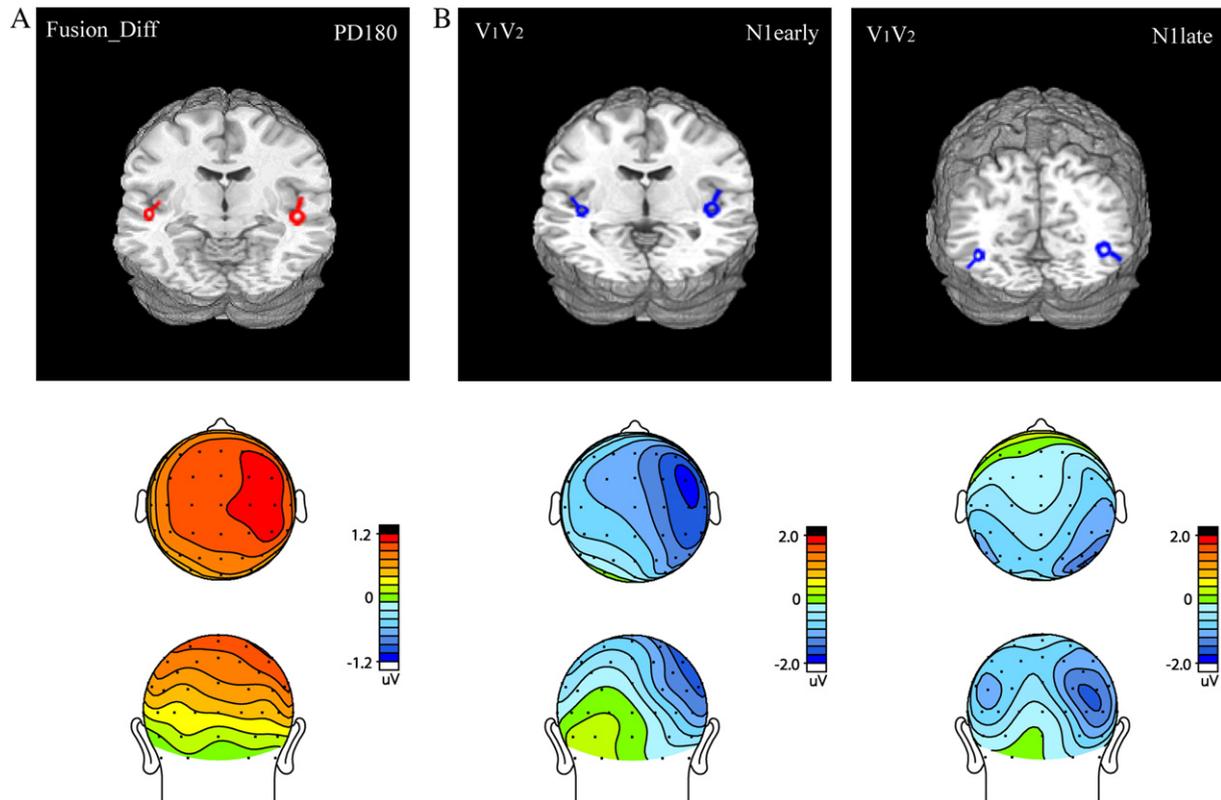

Fig. 5 – Estimated dipolar sources modeled using BESA and corresponding voltage topographies of the ERP components related to flash fusion in the SEE2 subject group. [A] Source model and topography of the PD180 component in the Fusion_Diff waveform. [B] Source models and topographies of the early (148–168 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.

range however, the amplitudes of the ERPs to $A_1V_1V_2$ did not differ significantly between the two subject groups (Table 3, $F(1,32)=0.14$, $p=$ n.s.). With respect to behavior the correlations across subjects between the early/ late N1 amplitudes to V_1V_2 and the percent fusion responses to the V_1V_2 stimulus approached but did not reach significance (early N1: $r(32)=0.25$, $p=$ n.s.; late N1: $r(32)=0.20$, $p=$ 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.

2.4. Source analysis

The neural generators of the components 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. Talairach coordinates of the dipole pairs and an estimate of
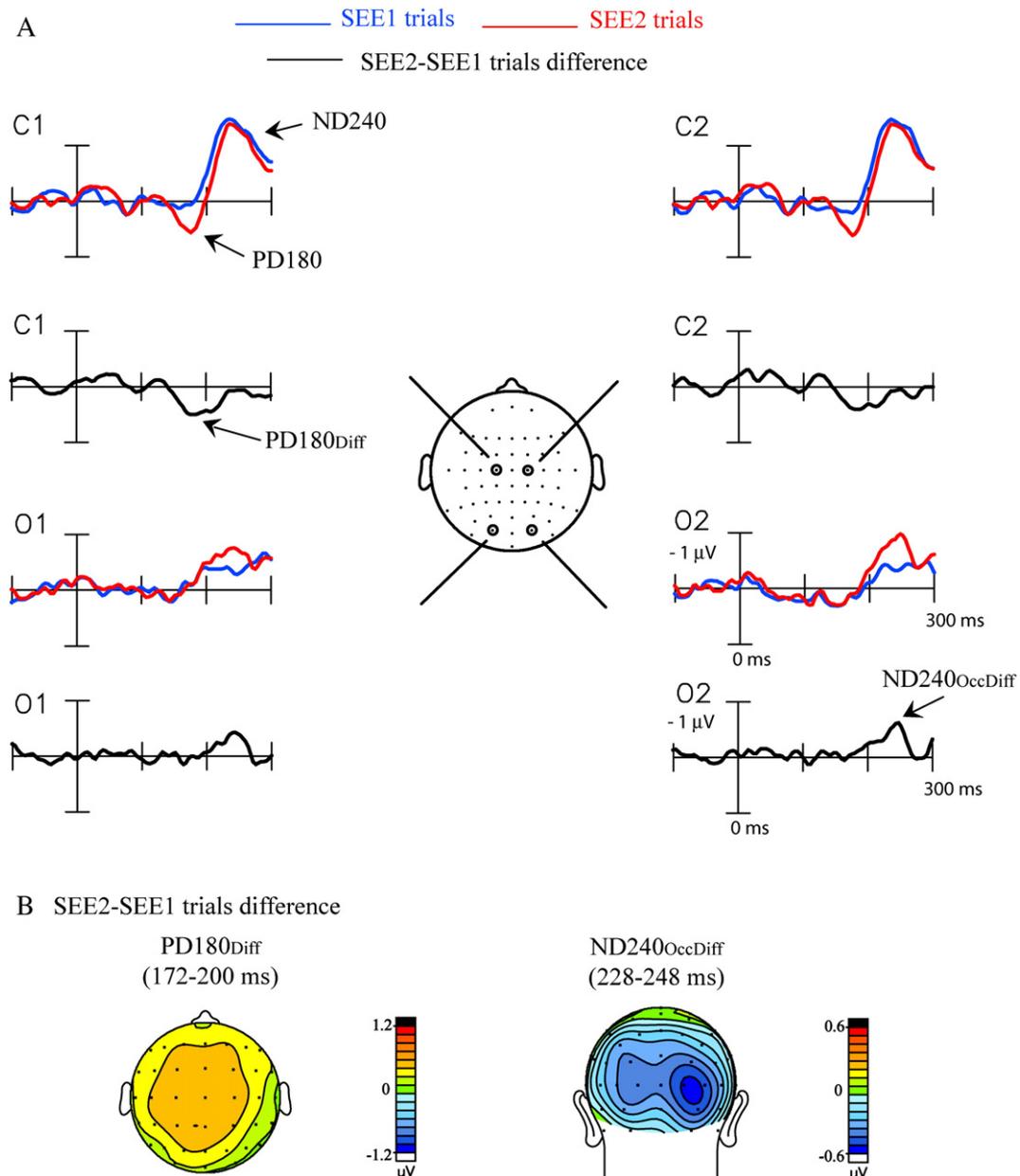

Fig. 6 – ERP differences between SEE1 and SEE2 trials for 15 subjects who had nearly equivalent numbers of SEE1/ SEE2 trials. [A] Fusion_Diff difference waves averaged separately for SEE1 and SEE2 trials. The SEE2–SEE1 trial difference wave reflects differential neural activity elicited on the SEE2 trials 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 difference wave.

their goodness of fit as reflected by residual variance are listed in Table 4.

The locations of the dipoles fit to the components that were correlated with perception of flash fusion (i.e., the PD180 component in the Fusion_Diff wave and the N1 component in the V_1V_2 ERP) are shown in Fig. 5. The PD180 component in the Fusion_Diff wave was localized to the region of the superior temporal gyrus bilaterally with a greater source amplitude in the right hemisphere. The later ND240 component in the Fusion_Diff wave was similarly localized to the superior temporal region (Table 4). The voltage distributions of the early (148–168 ms) and late (168–188 ms) phases of the visual N1 component elicited by V_1V_2 were modeled sequentially using a pair of dipoles in each phase (Fig. 5B). The source of the anterior N1 was localized to superior temporal cortex in close proximity to the PD180 dipoles in the Fusion_Diff waveform. The posterior N1 was 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, contralateral to the side of stimulus presentation.

2.5. 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 SEE1 and SEE2 trials revealed a significant difference within the PD180 latency range (172–200 ms) (SEE1 vs. SEE2 trials: $F(1,14) = 4.64$, $p < 0.05$, Fig. 6) with larger amplitude on SEE2 trials. A later trial difference in the ND240 time window (228–248 ms) was found to be significant over right occipital electrodes (SEE1 vs. SEE2 trials: $F(1,14) = 4.69$, $p < 0.05$). To distinguish this contralateral occipital effect from the previously described anterior ND240, it will be termed $ND240_{Occ}$. These trial specific differences were evident in the 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: $t(14) = -2.16$, $p < 0.05$).

It should be noted that the difference between the Fusion_Diff waveforms on SEE2 versus SEE1 trials is algebraically identical to the difference between the cross-modal ERPs elicited to $A_1V_1V_2$ on SEE2 versus SEE1 trials. This is because the ERPs to the unimodal stimuli that are subtracted to obtain the Fusion_Diff waveforms are identical for SEE2 and SEE1 trials. The SEE2–SEE1 trial difference was calculated on the Fusion_Diff waveforms in order to allow direct comparison with the Fusion_Diff waveforms described above for the SEE1 and SEE2 groups.

The voltage topography of the $PD180_{Diff}$ component was similar to that of the 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 × Electrode interaction: $F(37,518) = 1.26$, $p =$ n.s.). The later $ND240_{OccDiff}$ component had a topography centered over right visual cortex, which was significantly different from the topography of the centrally distributed ND240 component in the Fusion_Diff waveform ($ND240_{OccDiff}$ vs. ND240 × Electrode interaction: $F(37,518) = 6.45$, $p < 0.0001$) (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 4. The $PD180_{Diff}$ component was fit by dipole pairs with very similar coordinates as those of 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 dipoles in visual cortex with dipoles localizing to the lingual gyrus with a stronger right hemisphere source.

3. 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 = $[(A_1V_1V_2 + NoStim) - (A_1 + V_1V_2)]$. 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 the superior temporal cortex, which includes polysensory processing regions (Calvert et al., 2004). 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 veridical double flash percept is based on a greater cross-modal interaction within superior temporal cortex starting at around 100 ms after presentation of the second flash of the $A_1V_1V_2$ stimulus, which was 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). In any case, reduced amplitudes of the PD180 and $ND240_{OccDiff}$ components were strongly linked to the flash fusion percept.

The individual differences between subjects observed in the present study, especially with respect to their perceptual reports, can potentially explain why some previous studies 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 SEE1 and SEE2 groups 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 took place on the average but did 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 appears to 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 modulation found in the Fusion_Diff waveforms was the PD180 component (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 studies of cross-modal interactions (Teder-Sälejarvi et al., 2002, 2005; Molholm et al., 2002; Talsma and Woldorff 2005; Mishra et al., 2007), but this is the first report to our knowledge demonstrating its covariation with perception. In particular, variations in PD180 were not found to be 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 determined 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, especially its early phase (148–168 ms) 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-evoked N1s on V_1V_2 trials. The mean N1 amplitude difference between subjects who fused the double flash stimuli (SEE1 group) versus those who did not (SEE2 group) in the latency range of the PD180 component was found to be almost equivalent to the mean group difference for the PD180 component itself. This suggests that the variation of the PD180 component across subjects could largely be accounted

for by differences in the visual evoked N1 for these subjects. Indeed, the anteriorly distributed early phase of the visual N1 was found to have neural generators in close proximity to the PD180 source in superior temporal cortex. However, the larger N1 that was subtracted in the Fusion_Diff waveform in the SEE2 group cannot account for all the PD180 difference, because ERPs to $A_1V_1V_2$ had the same (positive) amplitude in the two groups within the PD180 time window. This indicates greater cross-modal interaction in the $A_1V_1V_2$ waveform in the SEE2 group in which the presence of A_1 reduced the larger N1 evoked to V_1V_2 such that $A_1V_1V_2$ amplitude in the N1/ PD180 latency range was equivalent in the two groups.

These results suggest that the neural basis of the flash fusion effect for both V_1V_2 and $A_1V_1V_2$ stimuli may involve sensory processing reflected in the N1 within the same superior temporal region. The early phase of the visual N1 has been reported to have multiple generators, both in temporal (Clark and Hillyard, 1996) and in parietal cortex (Di Russo et al., 2002, 2003). 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 these groups were found to show differential cross-modal interaction effects in auditory/ visual sensory cortices depending on which of their modalities was behaviorally dominant. Our findings suggest that individual differences in visual discrimination ability can also arise from processing differences in the superior temporal region.

The trial-by-trial analysis of the ERPs in a group who saw flash fusion on about half the trials revealed diminished PD180 amplitudes on SEE1 vs. SEE2 trials. In contrast to the cross-modal Fusion_Diff wave, this SEE2–SEE1 trial difference did not receive any contribution from the unimodal (A_1 or V_1V_2) ERPs. Thus, the trial-by-trial difference in the PD180 latency range originated solely from differential processing of the $A_1V_1V_2$ stimulus on SEE2 vs. SEE1 trials and was not a consequence of subtracting a larger N1 amplitude in the Fusion_Diff wave, which clearly contributed to the larger PD180 in the SEE2 vs. SEE1 groups as described above. In a later time window (228– 248 ms) a SEE1 vs. SEE2 trial difference was also found in visual cortex ($ND240_{OccDiff}$) that localized to ventral extrastriate areas near the fusiform gyrus. This component was unique to the SEE2–SEE1 trial difference and differed in topography and source localization from the ND240 component in the Fusion_Diff wave, whose neural generators lay in the vicinity of superior temporal cortex. Since the $ND240_{OccDiff}$ 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 entirely 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 their 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 with a reduced ability to discriminate the two flashes of $A_1V_1V_2$ as being separate, the large cross-modal interaction component PD180, onsetting 80–112 ms after V_2 and localizing to superior temporal area, was greatly diminished. Within these subjects the early phase of the anteriorly distributed N1 component to V_1V_2 stimuli (148–168 ms) was also significantly reduced. This early N1 was localized to the same superior temporal region as the PD180, while the later phase (168–188 ms) that was localized to extrastriate visual cortex did not show any group difference. The covariation of the PD180 and N1 amplitudes across subjects suggested that individual differences in perception of the cross-modal flash fusion phenomenon are driven in large part by individual differences in visual processing. A modulation in the PD180 latency range localized to the superior temporal area was also consistently observed in the trial-by-trial analysis of the ERPs, followed by a delayed modulation in extrastriate visual cortex (228–248 ms). These trial specific modulations were attenuated when the second flash was not perceived by subjects. Overall, these neural processes associated with flash fusion were found to be very different in their spatio-temporal pattern from the neural correlates of the sound-induced extra flash illusion (Mishra et al., 2007). The illusory extra flash generated to the $A_1V_1A_2$ stimulus was found to depend on an early sequence of activity (90–150 ms post-stimulus onset) involving auditory, visual and superior temporal cortices, all of which occurred before the emergence of the first cortical modulation associated with the flash fusion percept (the PD180). Hence, although the extra flash illusion and flash fusion may appear to be reciprocal phenomena their neural counterparts are very different. The present 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 an enhanced cross-modal interaction in superior temporal cortex. Activation of this multisensory region and subsequent feedback to visual cortex may enable accurate judgments of the timing and sequencing of visual stimuli in both unimodal and crossmodal contexts.

4.   Experimental procedures

4.1.   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.

4.2.   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 20° of visual angle to the left of fixation (Fig. 7A). Each visual stimulus was a 5 ms 75 cd/m^2 flash, and each auditory stimulus was a 10 ms 76 dB noise burst. Ten different stimulus combinations were presented in random order on each block of trials (Fig. 7B). 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 all stimulus combinations (except no-stim) is illustrated in Fig. 7. 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 ten 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
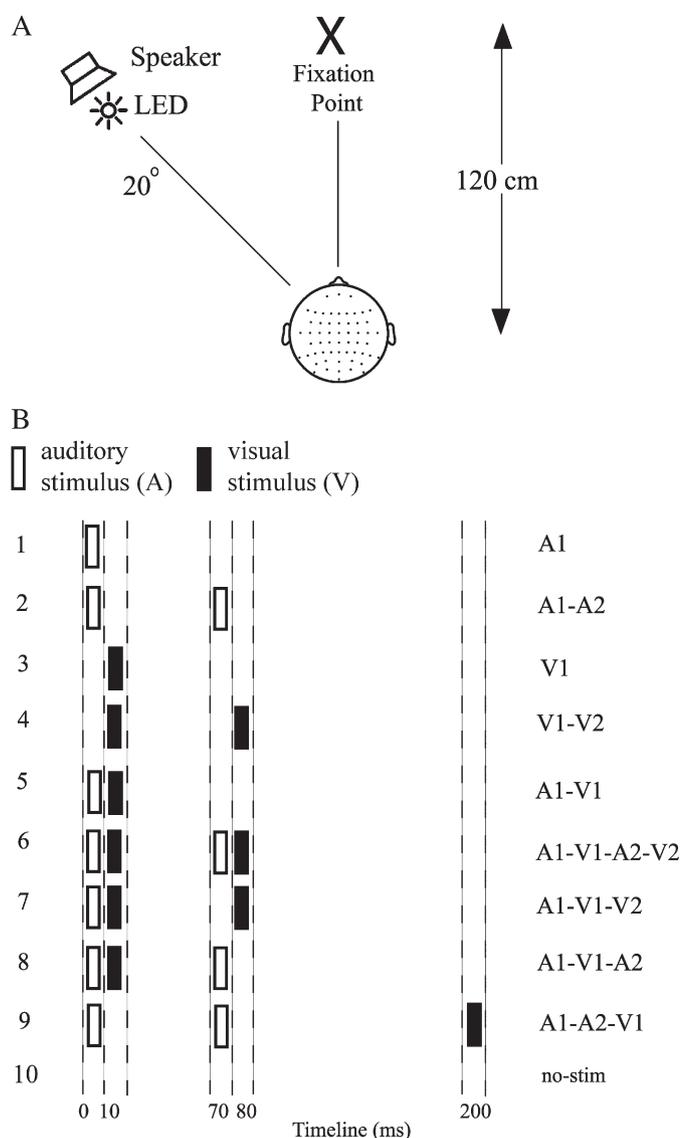
Fig. 7 – Overview of experimental design. [A] Schematic diagram of experimental set-up. [B] Listing of the ten different stimulus configurations, which were presented in random order. Abscissa indicates times of occurrence of auditory (open bars) and visual (solid bars) stimuli. Auditory (A) and visual (V) stimuli are labeled 1 or 2 to designate their first or second occurrence in each configuration (adapted from Mishra et al., 2007).

or more flashes. No responses were required to the unimodal auditory stimulation.

4.3. Behavioral analysis

For each stimulus configuration that contained one or two flashes, the percentages of one and two flash reports were calculated for each individual subject. Responses were scored within a 200–1200 ms period post-stimulus onset, and mean reaction times (RTs) were also calculated separately for each response type and stimulus. Percent responses as well as RTs were compared across stimulus conditions using t-tests. Given

the variability among subjects in percent fusion responses to the stimulus of interest, $A_1V_1V_2$, behavioral measures were also compared between subjects. For this analysis, the pool of 34 subjects was divided into two groups, designated SEE1 and SEE2 (17 in each), by a median split of the percent correct responses on the $A_1V_1V_2$ stimulus. The SEE1 group was the group of subjects that reported seeing one flash (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. 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).

In order to verify that differences in behavioral responses between the SEE1 and SEE2 subject groups were due to differences in perceptual sensitivity rather than decision bias, a signal detection analysis was performed (Macmillan and Creelman, 1991). For the $A_1V_1V_2$ stimulus, the average perceptual sensitivity estimate (d′) and the likelihood ratio criterion bias (ß) were calculated in each group. For each subject, correct two-flash responses to $A_1V_1V_2$ were categorized as "hits" and one-flash responses as "misses"; incorrect two-flash responses to A_1V_1 were considered "false alarms" and one-flash responses as "correct rejections". These d′ and ß estimates were compared between the SEE1 and SEE2 groups using t-tests. These signal detection parameters were also compared between the two subject groups for the V_1V_2 stimulus; in this case accurate two-flash responses to V_1V_2 were categorized as hits and incorrect two-flash responses to V_1 as false alarms.

4.4. Electrophysiological (ERP) recordings

The EEG was recorded from 62 electrode sites using a modified 10-10 system montage (Teder-Sälejarvi 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. The importance of fixation was emphasized to subjects, and the experimenter continually monitored the EOG and verified fixation in all blocks. 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; 3 dB 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 rereferenced 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: $Fusion\_Diff = [(A_1V_1V_2) + no\text{-}stim] - [A_1 + V_1V_2]$

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-Salejarvi et al., 2002; Talsma and Woldorff 2005; Gondan and Roder 2006; Mishra et al., 2007).

4.5. 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. The between-subject analysis was performed on the SEE1 and SEE2 subject groups described in the Behavioral analysis section above.

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.

4.6. Modeling of ERP sources

Source localization was carried out to estimate the intracranial generators of components in the grand-averaged ERPs and 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.

4.7. Trial based analysis

A trial-by-trial analysis of the ERPs elicited associated 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. Fusion_Diff 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 non-equivalent 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 $ND240_{Occ}$. PD180 differences between SEE2 and SEE1 trials were quantified as the mean voltage over the same fronto-central electrode clusters as specified above. The $ND240_{Occ}$ trial differences were measured over occipital sites (6 lateral electrodes in each hemisphere) where the differences were maximal.

REFERENCES

Andersen, T.S., Tiippana, K., Sams, M., 2004. Factors influencing audiovisual fission and fusion illusions. Brain Res. Cogn. Brain Res. 21, 301–308.

Arden, G.B., Wolf, J.E., 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. Vis. Res. 43, 2469–2478.

Bavelier, D., Neville, H.J., 2002. Cross-modal plasticity: where and how? Nat. Rev., Neurosci. 3, 443–452.

Calvert, G.A., Stein, B.E., Spence, C., 2004. The handbook of multisensory processing. MIT, Cambridge, MA.

Clark, V.P., Hillyard, S.A., 1996. Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. J. Cogn. Neurosci. 8, 387–402.

Coles, M.G.H., Smid, H.G.O.M., Scheffers, M.K., Otten, L.J., 1995. Mental chronometry and the study of human information processing. In: Rugg, M.D., Coles, M.G.H. (Eds.), Electrophysiology of mind. Oxford University Press, New York, pp. 86–131.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Di Russo, F., Martinez, A., Sereno, M.I., Pitzalis, S., Hillyard, S.A., 2002. Cortical sources of the early components of the visual evoked potential. Hum. Brain Mapp. 15, 95–111.

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

Fendrich, R., Corballis, P.M., 2001. The temporal cross-capture of audition and vision. Percept. Psychophys. 63, 719–725.

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

Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? Trends Cogn. Sci. 10, 278–285.

Giard, M.H., 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.

Hillyard, S.A., Picton, T.W., 1987. Electrophysiology of cognition. In: Plum, F. (Ed.), Handbook of Physiology Section 1: The Nervous System. Higher Functions of the Brain, Part 2, Vol. V. American Physiological Society, Bethesda, Maryland, pp. 519–584.

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

Macmillan, N.A., Creelman, C.D., 1991. Detection theory: A user's guide. Cambridge University Press, New York.

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

McDonald, J.J., Teder-Salejarvi, W.A., Di Russo, F., Hillyard, S.A., 2003. Neural substrates of perceptual enhancement by cross-modal spatial attention. J. Cogn. Neurosci. 15, 10–19.

McDonald, J.J., Teder-Salejarvi, W.A., Di Russo, F., Hillyard, S.A., 2005. Neural basis of auditory-induced shifts in visual time-order perception. Nat. Neurosci. 8, 1197–1202.

Meylan, R.V., Murray, M.M., 2007. Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. Neuroimage 35, 244–254.

Mishra, J., Martinez, A., Sejnowski, T.J., Hillyard, S.A., 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, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res. Cogn. Brain Res. 14, 115–128.

Noesselt, T., Bonath, B., Boehler, C.N., Schoenfeld, M.A., Heinze, H.J., 2007a. On perceived synchrony-neural dynamics of audiovisual illusions and suppressions. Brain Res., doi:10.1016/j.brainres.2007.09.045.

Noesselt, T., Rieger, J.W., Schoenfeld, M.A., Kanowski, M., Hinrichs, H., Heinze, H.J., Driver, J., 2007b. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J. Neurosci. 27, 11431–11441.

Recanzone, G.H., 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. In: Grandori, F., Hoke, M., Roman, G.L. (Eds.), Auditory evoked magnetic fields and electric potentials, pp. 40–69.

Schroeder, C.E., Foxe, J., 2005. Multisensory contributions to low-level, 'unisensory' processing. Curr. Opin. Neurobiol. 15, 454–458.

Sekuler, R., Sekuler, A.B., 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, W.J., Beierholm, U., 2005b. Sound-induced flash illusion as an optimal percept. Neuroreport 16, 1923–1927.

Stein, B.E., London, R., Wilkinson, L.K., Price, D.D., 1996. Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J. Cogn. Neurosci. 8, 497–506.

Stein, B.E., Meredith, M.A., 1993. The merging of the senses. MIT, Cambridge, MA.

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

Talsma, D., Woldorff, M.G., 2005. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J. Cogn. Neurosci. 17, 1098–1114.

Teder-Salejarvi, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A., 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, W.A., Di Russo, F., McDonald, J.J., Hillyard, S.A., 2005. Effects of spatial congruity on audio-visual multimodal integration. J. Cogn. Neurosci. 17, 1396–1409.

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, J.D., Rees, G., 2006. Sound alters activity in human V1 in association with illusory visual perception. Neuroimage 31, 1247–1256.