Behavioral/Systems/Cognitive

Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion

Jyoti Mishra,[1] **Antigona Martinez,**[2,3] **Terrence J. Sejnowski,**[1,4] **and Steven A. Hillyard**[2]

[1]Division of Biological Sciences and [2]Department of Neurosciences, University of California, San Diego, La Jolla, California 92093, [3]Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York 10962, and [4]Howard Hughes Medical Institute, Computational Neurobiology Laboratory, Salk Institute, La Jolla, California 92037

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.

Key words: ERPs; auditory cortex; visual cortex; illusory flash; cross-modal interaction; source analysis

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 that 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 at both very early and late time periods after stimulus onset (Giard and Peronnet, 1999; Molholm et al., 2002; Murray et al., 2005; Meylan and Murray, 2007; Talsma et al., 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; Fendrich and Corballis, 2001; McDonald et al., 2003, 2005; Recanzone, 2003; Vroomen and de Gelder, 2004). A particularly striking perceptual alteration was recently described by Shams et al. (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, because 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.

Previous ERP/magnetoencephalographic (Shams et al., 2001, 2005a; Arden et al., 2003) and functional magnetic resonance (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
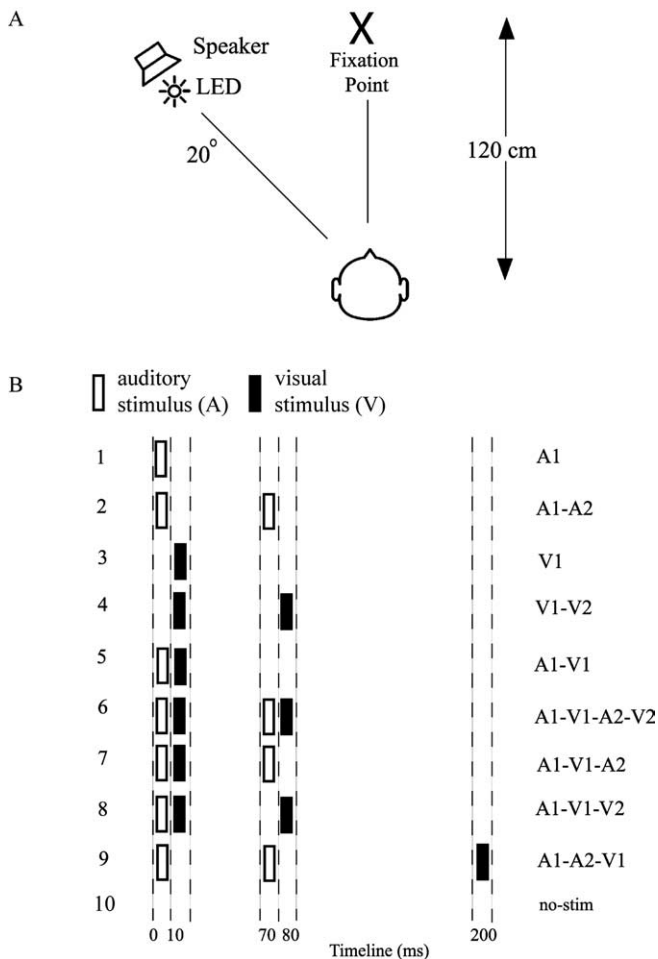
Figure 1. Overview of experimental design. *A*, Schematic diagram of experimental setup. *B*, Listing of the 10 different stimulus configurations, which were presented in random order. Abscissa indicates times of occurrence of auditory (open bars) and visual (solid bars) stimuli. Auditory (A) and visual (V) stimuli are labeled 1 or 2 to designate their first or second occurrence in each configuration. LED, Light-emitting diode.

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 spatiotemporal 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 and Methods

Task and stimuli. Thirty-four right-handed healthy adults (18 females; mean age of 23.9 years) 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, respectively, both positioned 20° of visual angle to the left of fixation (Fig. 1 *A*). 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. 1 *B*). 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, $A_1V_1A_2V_2$, $A_1V_1A_2$, $A_1V_1V_2$ 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) trial 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 Figure 1 *B*. The stimulus onset asynchrony (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 $A_1A_2V_1$, in which 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 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; 3 dB attenuation) and were digitized at 250 Hz. Automated artifact rejection was performed before averaging to discard trials with eye movements, blinks, or amplifier blocking. Signals were averaged in 500 ms epochs with a 100 ms prestimulus 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

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	Percentage of trials reporting two flashes over all subjects (SEE group/NO-SEE group)	SEM (% trials) over all subjects (SEE group/NO-SEE group)	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

Percentage trials on which two flashes were reported and the SEM of these percentages are reported over all 34 subjects and separately (in parentheses) for the SEE and NO-SEE subject groups (n = 17 each). RT, Reaction time.

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 (Colchester, VT) spatial digitizer. 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: Ill_Diff = $[(A_1V_1A_2) + \text{no-stim}] - [A_1A_2 + V_1]$; (2) neural activity to real second flash: 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 = $[(A_1V_1 + \text{no-stim})] - [A_1 + V_1]$; (2) A_2V_2_Diff = $[(A_1V_1A_2V_2 + \text{no-stim})] - [A_1A_2 + V_1V_2]$.

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 contingent negative variation that may extend into the poststimulus 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 and Woldorff, 2005; Gondan and 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: eight males, nine females, mean age of 24 years; NO-SEE group: eight males, nine females, mean age of 23.8 years). 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 below, Modeling of ERP sources). Finally, a trial-by-trial analysis was performed in which ERPs and cortical oscillations were compared for trials on which the illusion was perceived ("SEE" trials) versus not perceived ("NO-SEE" trials) (see below, Frequency domain analysis).

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 at which its amplitude was maximal. The PD120 and ND160 components were averaged over nine occipital electrode sites spanning the midline, PD180 amplitude was measured over frontocentral electrode clusters (eight in each hemisphere and four over midline), the ND270 and ND275 were measured over central electrode sites (eight in each hemisphere and four over midline), and the PD200 was averaged over eight occipital electrode sites in the right hemisphere (contralateral to side of stimulation).

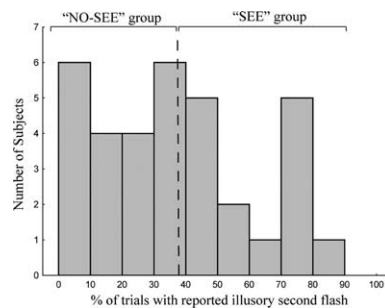Scalp distributions of these ERP components in the Ill_Diff and Vis-

Figure 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).

_Diff difference waves were compared after normalizing their amplitudes before 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 (seven in each hemisphere and four 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 eight 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 performed 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 three-dimensional current density distributions using a realistic head model with a solution space of 4024 nodes equally distributed within the gray matter of the average template brain of the Montreal Neurological Institute. 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 groups. 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-averaged 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 frontocentral electrode clusters as those used to measure PD180 in the Ill_Diff waveform (see above, Data analysis).
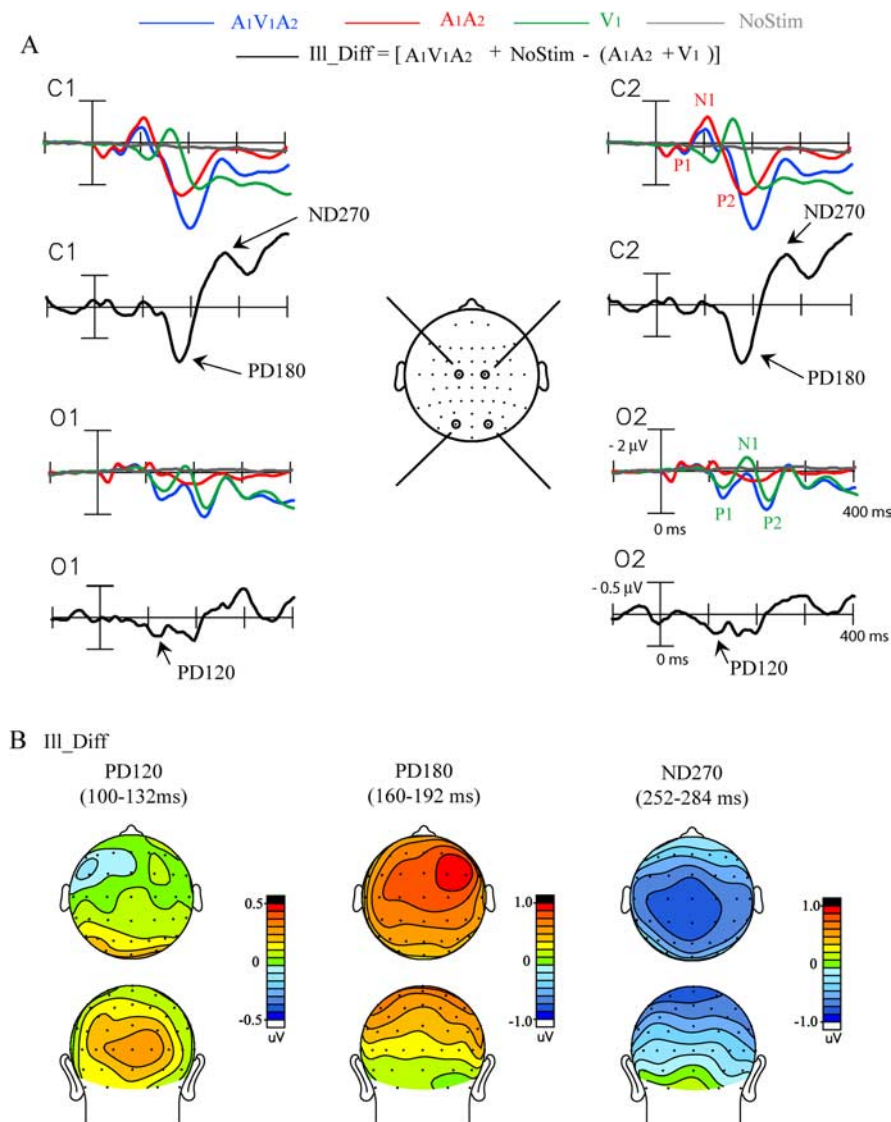
Figure 3. Grand-averaged ERPs ($n = 34$) associated with the sound-induced illusory flash. **A**, ERPs elicited by the illusion-inducing $A_1V_1A_2$ stimulus and by its unimodal constituents A_1A_2 and V_1, together with the ERP time locked to the blank no-stim event. The Ill_Diff difference wave (see Materials and Methods) reflects the cross-modal interactions giving rise to the illusory second flash. Recordings are from left and right central (C1, C2) and occipital (O1, O2) sites. **B**, Topographical voltage maps of the three major components in the Ill_Diff difference wave shown in top and back views.

quency scale. This analysis revealed significant differences within the 20–50 Hz frequency range over the occipital scalp. Because multiple pointwise 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 (six in each hemisphere) in the 25–35 Hz frequency range and 204–236 ms over 16 occipitoparietal sites (eight 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 $A_1A_2V_1$, $F_{(1,33)} = 52.98$, $p < 0.0001$), although 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 reported previously 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 <10% to >80% (Fig. 2). 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 percentage correct performance for any of the other stimuli (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 = $ NS) or for any of the other stimuli.

ERP results

Figure 3A shows the grand-averaged ERPs (over all 34 subjects) elicited by the illusion-inducing $A_1V_1A_2$ stimulus and by its uni-

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 s 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 microvolts). 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 prestimulus 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 because of the 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 fre-

modal 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 visually 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 Materials and Methods) are also shown in Figure 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 frontocentral sites with a nonsignificant 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 centroparietal 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 Figure 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 Materials and 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 0 (defined as 10 ms before V_1 onset) (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, 120, and 195 ms, respectively, after presentation of the second flash. This timing suggests an equivalence with the well known visually 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 wave-

Table 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)

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

Components were measured over scalp sites of maximal amplitude, as described in Materials and Methods. Significance levels of component amplitudes were tested with respect to the 100 ms prestimulus baseline.
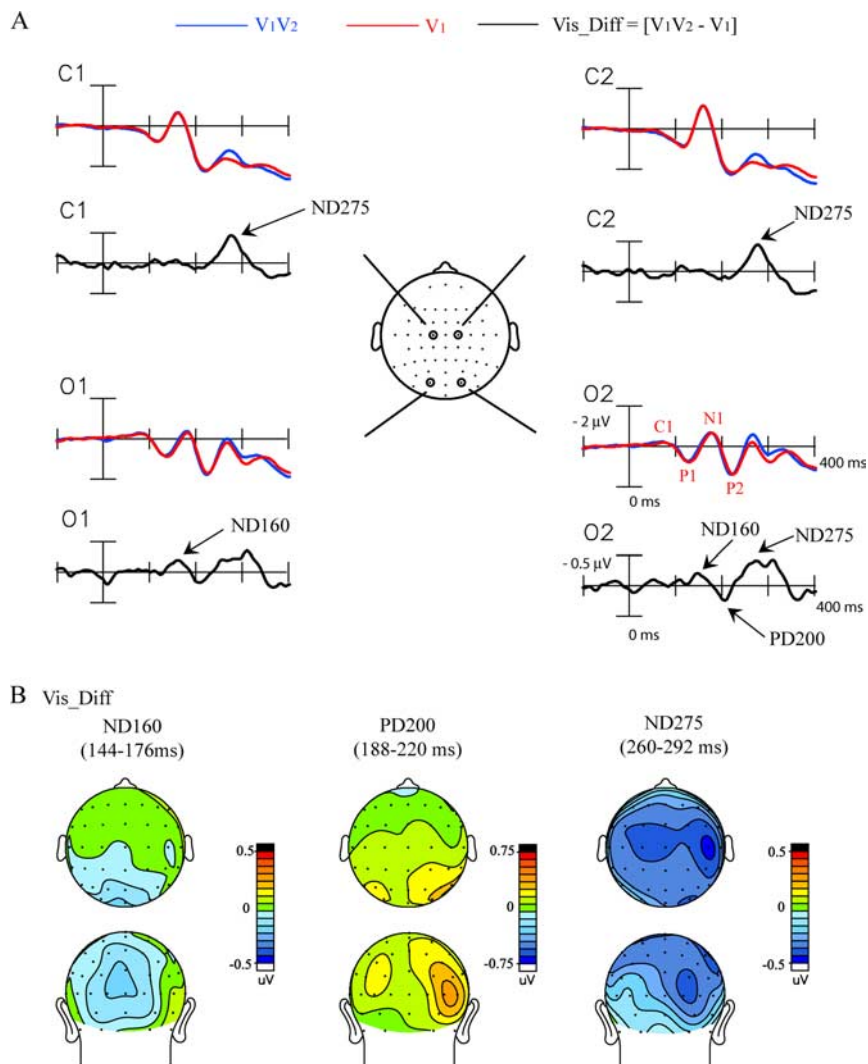

Figure 4. Grand-averaged ERPs ($n = 34$) associated with the veridical second flash. *A*, ERPs elicited by the pair of flashes, V_1V_2, and by the single flash, V_1. The Vis_Diff difference wave reflects neural activity elicited by the second flash, V_2. Recordings are from left and right central (C1, C2) and occipital (O1, O2) sites. *B*, Topographical voltage maps of the three major components in the Vis_Diff difference wave shown in top and back views.
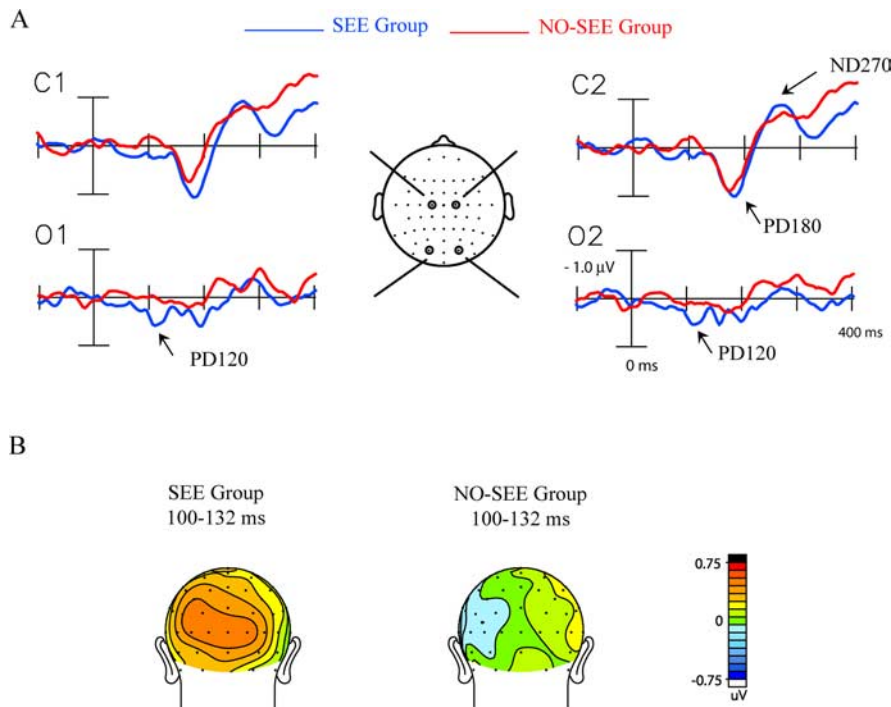
Figure 5. ERP differences between the SEE and NO-SEE groups. ***A***, Ill_Diff difference waves averaged separately for the SEE group (n = 17) and the NO-SEE group (n = 17). Recordings are from left and right central (C1, C2) and occipital (O1, O2) sites. ***B***, Voltage maps in back view comparing the topography of the PD120 component in the Ill_Diff difference waves in the two groups.

forms were compared after normalization according to the method of McCarthy and Wood (1985). The topography of the PD120 component in the Ill_Diff waveform differed significantly from the topographies of both the ND160 (condition by electrode interaction, $F_{(17,561)} = 3.02$, $p < 0.0001$) and the PD200 (condition by 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 by electrode interaction, $F_{(37,1221)} = 7.47$, $p < 0.0001$). Last, the Ill_Diff ND270 topography was significantly different from that of the Vis_Diff ND275 (condition by 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

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 fur-

ther 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 ($r = 0.03$; $p =$ NS) or ND270 ($r = -0.17$; $p =$ NS) 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 Materials and Methods). A between-group comparison of the Ill_Diff and A_2V_2_Diff waveforms over occipital site Oz is shown in Figure 6. For the SEE group, PD120 (measured over 100–132 ms) was significant with respect to baseline in both the Ill_Diff ($t_{(16)} = 4.18$; $p < 0.0008$) 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 by electrode interaction, $F_{(17,272)} = 0.72$, $p =$ NS) (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 summary, 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

Table 3. Comparison of component amplitudes in the Ill_Diff waveform between the SEE and NO-SEE subject groups

ERP component	Ill_Diff (SEE group)		Ill_Diff (NO-SEE group)		$F_{(1,32)}$	$p <$
	Amplitude (μV)	SEM (μV)	Amplitude (μV)	SEM (μV)		
PD120	0.43	0.10	0.02	0.06	11.42	0.002
PD180	0.85	0.23	0.63	0.21	0.49	NS
ND270	−0.73	0.21	−0.69	0.20	0.02	NS

Components were measured over scalp sites of maximal amplitude, as described in Materials and Methods.
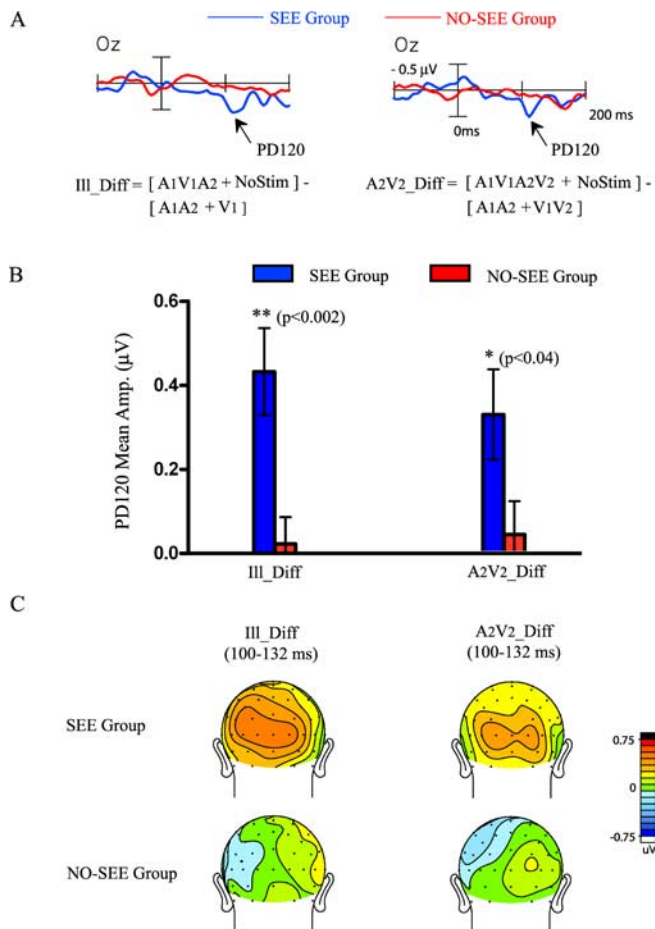
Figure 6. Comparison of the PD120 component elicited in the Ill_Diff and A$_2$V$_2$_Diff cross-modal difference waves for the SEE and NO-SEE groups. **A**, Waveforms of the Ill_Diff and A$_2$V$_2$_Diff difference waves for the two groups recorded from an occipital electrode (Oz). **B**, Bar graphs comparing the mean amplitude of PD120 in the interval 100–132 ms in the Ill_Diff and A$_2$V$_2$_Diff waveforms for the two groups. **C**, Voltage maps comparing PD120 topographies for the two groups.

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$_1$V$_2$, 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, because 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).

Because the PD120 was the main ERP component differentiating the SEE and NO-SEE subject groups (see above, Between-subject analysis), an additional statistical analysis was performed 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 performed, 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 ± 2.8 total trials (mean ± SEM) 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 versus 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 and 124–156 ms after 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 = $ NS; 124–156 ms, $F_{(1,16)} = 0.17$, $p = $ NS). These differences between the groups were also evident in an overall ANOVA (with subject group as factor) as a group by 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 attributable 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 prestimulus 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 was nonsignificant. Both the ND110 and ND130 components
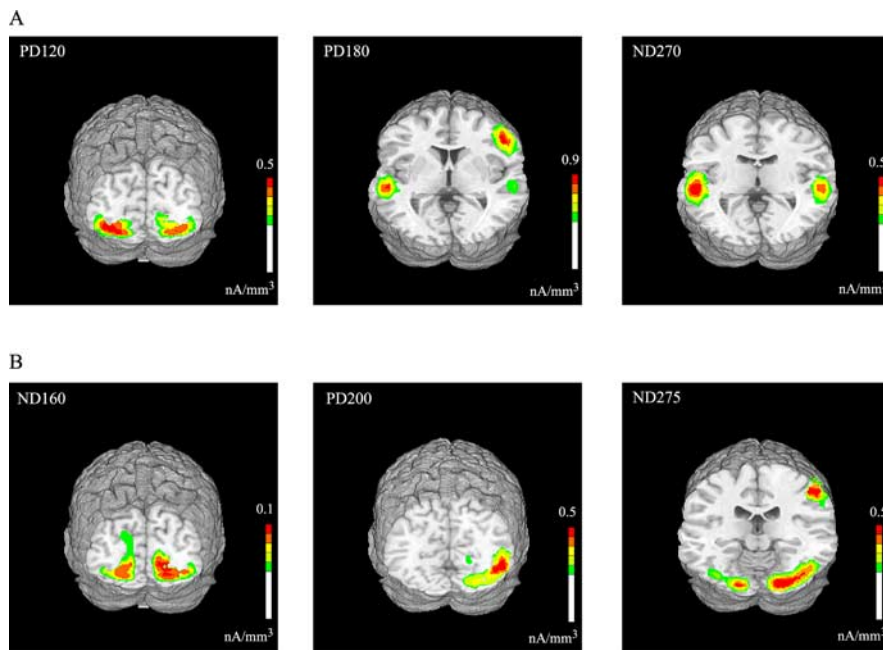
A

B

Figure 7. Estimated sources for the major components in the grand-averaged Ill_Diff (*A*) and Vis_Diff (*B*) 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 (nanoamperes per cubic millimeter).

had amplitude maxima at frontocentral sites (Fig. 8*B*). The right hemispheric preponderance of the ND130 component in the grand-averaged 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.

Because 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 with 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 = $ NS), 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 (Fig. 10*A*, post-auricular reflex). 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 110–145 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 by electrode, $F_{(5,80)} = 5.53$, $p < 0.0003$). The spatial topography of EP130 is shown in Figure 10*B*. For individual channels at which the EP130 was maximal, the SEE versus 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 = $ NS), which indicates that the EP130 effect is attributable 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

Table 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

$(F_{(1,16)} = 4.56; p < 0.05)$. Over individual occipitoparietal 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 O$_2$ $(t_{(16)} = 2.35; p < 0.033)$. 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-averaged 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 $(F_{(1,16)} = 1.92; p = NS)$ was found in the spectral amplitude within the -300 to -50 ms interval that was used as prestimulus 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 + $ no-stim$) - (A_1A_2 + V_1)]$. 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 in which 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 nonspecific areas as summarized in Figure 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 frontocentral scalp sites and was
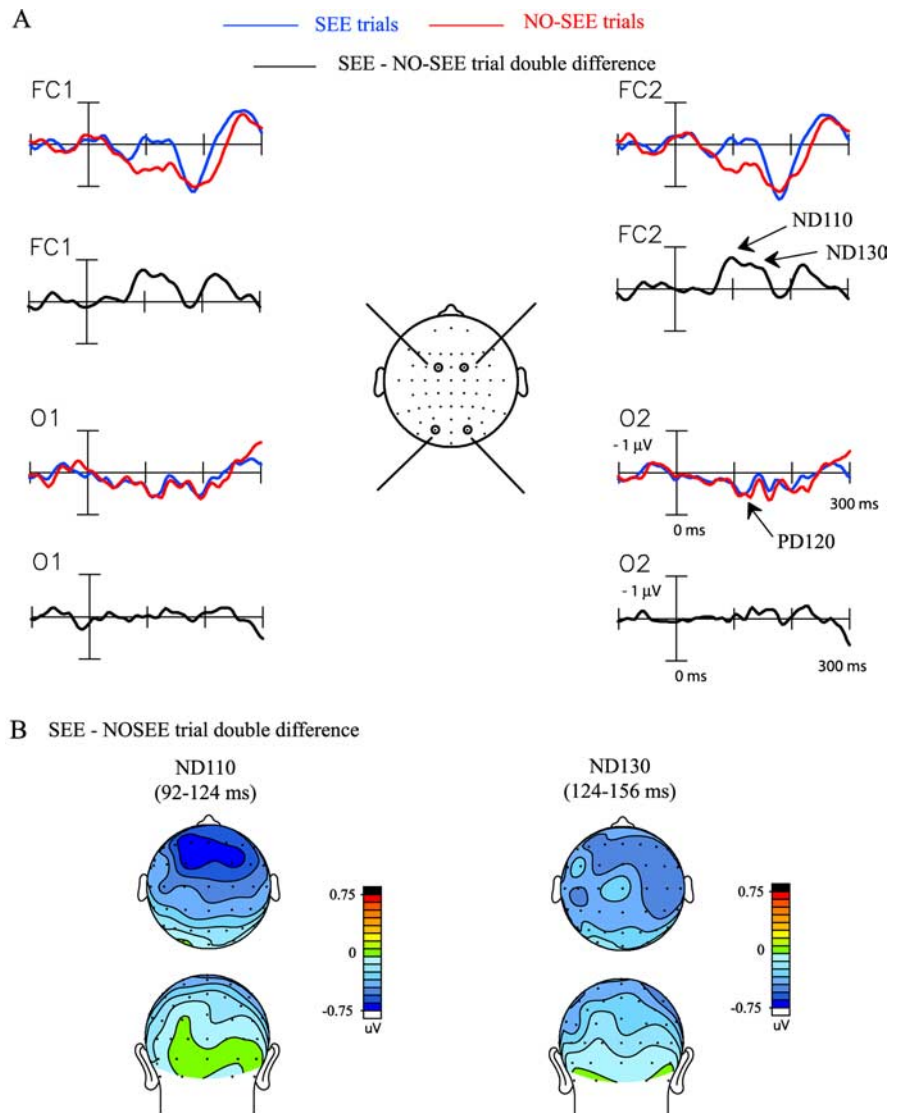

Figure 8. ERP differences between SEE and NO-SEE trials within the SEE subject group. **A**, Ill_Diff difference waves within the SEE group averaged separately for SEE and NO-SEE trials. The SEE–NO-SEE trial double-difference wave reflects differential neural activity elicited on the SEE trials with respect to NO-SEE trials. Recordings are from left and right frontocentral (FC1, FC2) and occipital (O1, O2) 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.
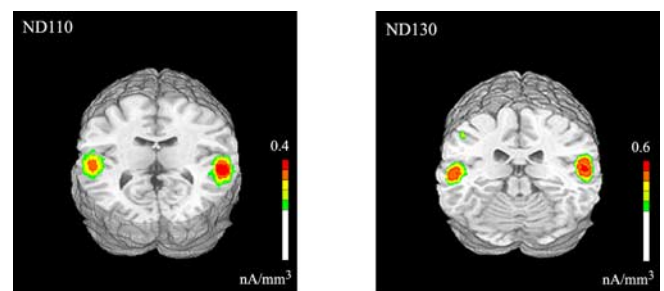

Figure 9. Estimated sources for the two early components in the SEE–NO-SEE trial double-difference wave modeled in the SEE group using LAURA. Results are shown on a standard fMRI rendered brain in Talairach space. LAURA inverse solutions are represented in units of current source density (nanoamperes per cubic millimeter).

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 tem-
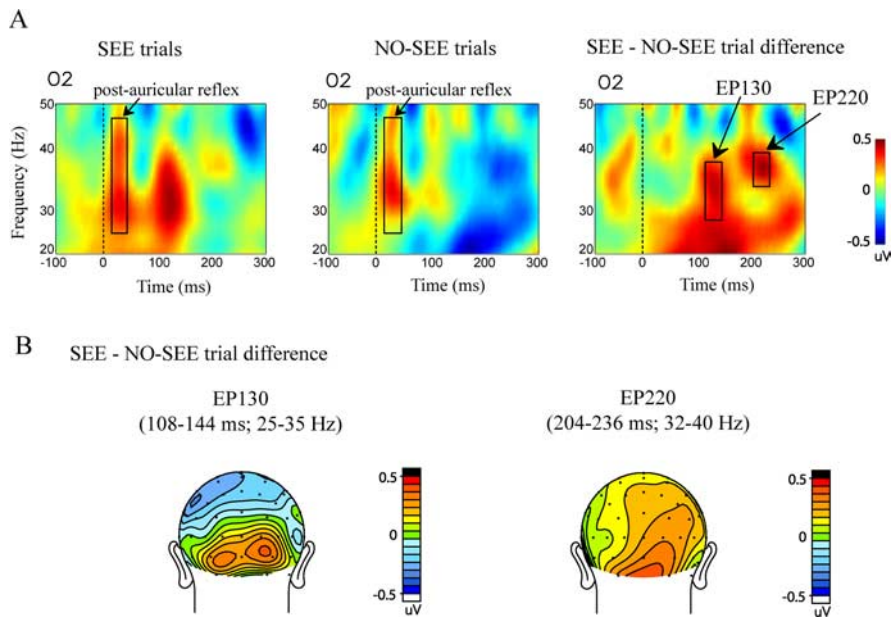
Figure 10. Frequency domain activity associated with perception of the sound induced illusory flash in the SEE subject group. *A*, Time–frequency representation of the total average spectral amplitude on SEE trials, NO-SEE trials, and the SEE–NO-SEE trial difference from an occipital electrode (O2). *B*, Spatial topography maps of the two time–frequency blocks of differential spectral amplitude, EP130 and EP220, found in the SEE–NO-SEE trial difference shown in back view.
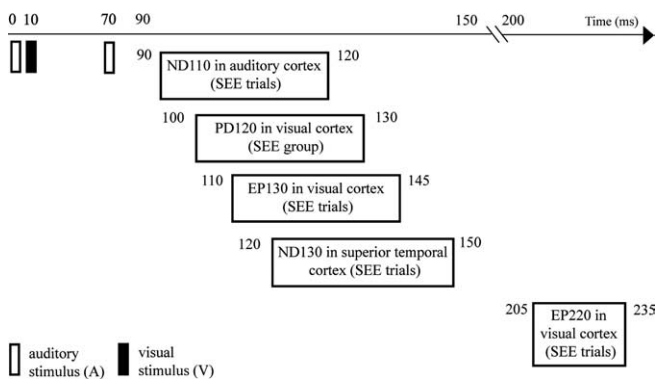

Figure 11. Summary of temporal progression of early cortical activity found to be associated with the sound induced extra flash illusion.

poral cortex, which is a well documented site of cross-modal interaction (for review, see 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 A_2V_2_Diff) suggests that they reflect general aspects of cross-modal interaction unrelated to perception of an illusory extra flash (Molholm et al., 2002; Teder-Sälejärvi et al., 2002, 2005; Talsma and 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 as follows: 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 occipitoparietal 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 (Clark and Hillyard, 1996; Di Russo et al., 2002, 2003).

The ERP correlates of illusory and real flash perception have been compared previously (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 at ∼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 ∼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, because of 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 of how frequently an individual subject will perceive the illusion. Recent anatomical labeling studies in macaques (Falchier et al., 2002; Rockland and Ojima, 2003; Clavagnier et al., 2004) 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 (20° in the periphery) suggests that the audiovisual 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 versus 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 additional 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 et al. (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.

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 USA 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.

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, 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. The 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, in press.

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.