Effects of Spatial Congruity on Audio-Visual Multimodal Integration

W. A. Teder-Sälejärvi[1], F. Di Russo[2], J. J. McDonald[3], and S. A. Hillyard[1]

Abstract

■ Spatial constraints on multisensory integration of auditory (A) and visual (V) stimuli were investigated in humans using behavioral and electrophysiological measures. The aim was to find out whether cross-modal interactions between A and V stimuli depend on their spatial congruity, as has been found for multisensory neurons in animal studies (Stein & Meredith, 1993). Randomized sequences of unimodal (A or V) and simultaneous bimodal (AV) stimuli were presented to right- or left-field locations while subjects made speeded responses to infrequent targets of greater intensity that occurred in either or both modalities. Behavioral responses to the bimodal stimuli were faster and more accurate than to the unimodal stimuli for both same-location and different-location AV pairings. The neural basis of this cross-modal facilitation was studied by comparing event-related potentials (ERPs) to the bimodal AV stimuli with the summed ERPs to the unimodal A and V stimuli. These comparisons revealed neural interactions localized to the ventral occipito-temporal cortex (at 190 msec) and to the superior temporal cortical areas (at 260 msec) for both same- and different-location AV pairings. In contrast, ERP interactions that differed according to spatial congruity included a phase and amplitude modulation of visual-evoked activity localized to the ventral occipito-temporal cortex at 100–400 msec and an amplitude modulation of activity localized to the superior temporal region at 260–280 msec. These results demonstrate overlapping but distinctive patterns of multisensory integration for spatially congruent and incongruent AV stimuli. ■

INTRODUCTION

Many objects and events in the natural world have multimodal features that stimulate different sensory pathways. Despite the initial segregation of the senses, information about the different features of an object is integrated in the brain to form unified multisensory percepts. Studies of human performance have revealed numerous behavioral consequences of multisensory integration. Under many circumstances, for example, the simultaneous stimulation of two or more modalities has a facilitatory effect on behavioral performance, with bimodal stimuli being detected and discriminated faster and more accurately than the constituent unimodal stimuli presented alone (for reviews, see Frassinetti, Bolognini, & Làdavas, 2002; Frens, Van Opstal, & Van der Willigen, 1995; Welsh & Warren, 1986; Loveless, Brebner, & Hamilton, 1970).

In groundbreaking studies of the neural bases of multisensory integration, Stein (1998) and Stein and Meredith (1993) discovered neurons in the cat brain that were responsive to more than one modality. These multisensory neurons were found in the superior colliculus as well as in certain cortical association areas. Many of these cells were particularly responsive to concurrent bimodal stimulation when the constituent stimuli appeared in close spatial proximity. For such cells, the neural response to a bimodal stimulus was larger than the sum of the responses to the individual unimodal stimuli. This multisensory response enhancement parallels the aforementioned behavioral facilitation resulting from bimodal stimulation and was proposed as the neural basis for the integration of unimodal sensations into unified multimodal percepts (Stein & Meredith, 1993).

The neural bases of multisensory integration have been studied in humans using neuroimaging techniques (Beauchamp, Lee, Argall, & Martin, 2004; Foxe, Wylie, et al., 2002; Laurienti et al., 2002; Calvert, Hansen, Iversen, & Brammer, 2001; Calvert, Campbell, & Brammer, 2000) and recordings of event-related brain potentials (ERPs) (Fort, Delpuech, Pernier, & Giard, 2002; Molholm et al., 2002; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002; Foxe, Morocz, et al., 2000; Giard & Peronnet, 1999; Schröger & Widmann, 1998) and magnetic fields (Raij, Uutela, & Hari, 2000). In several of these ERP studies, auditory–visual integration was investigated by comparing the ERP to a bimodal audio-visual (AV) stimulus with the sum of the ERPs to the constituent auditory (A) and visual (V) stimuli. Multisensory interactions were revealed in the difference

[1]University of California—San Diego, [2]University Institute of Motor Sciences (IUSM) and Santa Lucia Foundation IRCCS, Rome, Italy, [3]Simon Fraser University, Vancouver, Canada

waveform formed by subtracting the sum of the ERPs to the individual A and V stimuli from the ERP to the bimodal AV stimulus. Interaction components were observed in this AV − (A + V) difference wave at short latencies (40–80 msec) over posterior scalp areas and at longer latencies (100–250 msec) over multiple scalp areas. The short latency effects were attributed by some authors (Molholm et al., 2002; Giard & Peronnet, 1999) to interactions between auditory and visual inputs at an early stage of processing in the visual cortex, but Teder-Sälejärvi et al. (2002) cautioned that such early effects might be contaminated by slow anticipatory potentials that are doubly subtracted in the AV − (A + V) difference wave. Scalp mapping and dipole modeling of the longer latency components in the difference waves suggested that auditory–visual interactions were taking place in ventral occipito-temporal, superior temporal, central–parietal, and frontal cortical regions (Fort et al., 2002; Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Giard & Peronnet, 1999).

In the neurophysiological studies of Stein and Meredith (1993), multisensory enhancement of neural responses (with AV > A + V) was observed primarily when the visual and auditory stimuli were presented in close spatial proximity. When the stimuli were widely separated in space, either response depression (AV < A + V) or no interaction (AV = A + V) was typically observed. An analogous beneficial effect of spatial congruity was seen in behavioral studies with cats (Stein, Meredith, Huneycutt, & McDade, 1989); the animals' ability to orient towards a dim visual cue was improved by the presence of a concurrent sound at the same location and was disrupted by a sound at a different location. In human studies, the facilitatory effect of bimodal AV stimulation on signal detectability and motor response speed was also found to depend on the spatial congruence of the stimuli (Frassinetti et al., 2002; Harrington & Peck, 1998; Corneil & Munoz, 1996; Frens et al., 1995; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; Bernstein & Edelstein, 1973; Simon & Craft, 1969). A psychophysical study by Stein, London, Wilkinson, and Price (1996) found a contrasting result, however; the brightness of a visual stimulus was found to be enhanced by a concurrent sound regardless whether the two stimuli were spatially coincident.

Only a few studies have investigated the neural mechanisms by which spatial congruity affects multisensory integration in humans. In a study with fMRI, Macaluso, Frith, and Driver (2000) found that light-evoked activity in the ventral visual cortex was enhanced when a spatially congruent tactile stimulus accompanied the light. Cross-modal interactions between lateralized visual (V) and somatosensory (S) were also observed in ERP recordings by Schürmann, Kolev, Menzel, and Yordanova (2002); of particular interest was a facilitatory interaction (VS > V + S) at a latency of 75–100 msec found only for spatially coincident stimulus pairs. In another ERP study, Murray et al. (2005) found facilitatory cross-modal interactions between lateralized auditory and somatosensory brain responses as early as 50–95 msec. This early AS interaction was localized to the auditory association cortex and was found to be equivalent for both spatially congruent and incongruent pairings. Thus, in line with the behavioral study of Stein, London, Wilkinson, et al. (1996), it appears that some types of facilitatory cross-modal interactions do not depend upon the spatial coincidence of the stimuli.

The role of spatial congruity on AV multisensory convergence was investigated with ERP recordings by Gondan, Niederhaus, Röder, and Rösler (in press). They observed an ERP interaction component that differed between spatially congruent and incongruent AV stimuli, which was maximal over the posterior parietal scalp at a latency of 150–180 msec. The present study extends this line of investigation using high-density ERP recordings and source localization techniques to analyze in more detail the neural bases of AV interactions that vary with spatial congruity. ERPs were recorded to randomized stimulus sequences that included individual A and V stimuli presented to left or right locations and simultaneous bimodal AV pairs presented either to the same location (left or right) or to opposite locations (left/right or right/left) (see Figure 1). Subjects were instructed to respond to infrequent targets of higher intensity that could occur in either modality (or both) at either location. Cross-modal interactions for spatially congruent

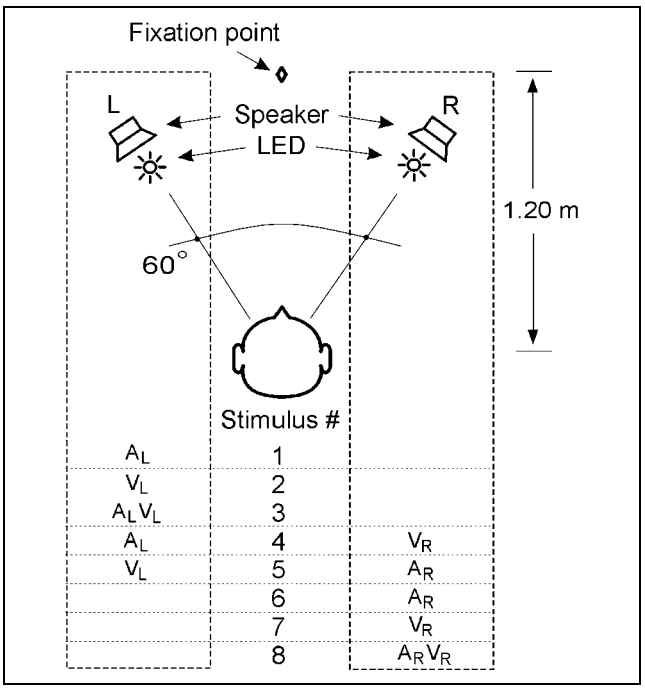

Figure 1. Schematic layout of left $_{(L)}$ and right $_{(R)}$ speakers and LEDs. The eight possible combinations of auditory (A), visual (V), and bimodal (AV) stimuli are listed below.

and spatially incongruent AV pairs were evident as components in the appropriate AV − (A + V) ERP difference waves. The timing and cortical localization of these cross-modal interactions were studied through dipole modeling of the interaction components in the ERP difference waveforms.

RESULTS

Behavioral Results

Detection Accuracy

Although target detectability in the two modalities was approximately equated during the practice session, in the actual experiment the unimodal visual targets were detected more accurately than the auditory targets [$t(14) = 5.11$, $p < .001$] (Table 1). Even with this imbalance, however, bimodal targets were correctly detected at a much higher rate overall than either the unimodal auditory [$t(14) = 5.92$, $p < .0001$] or visual [$t(14) = 3.58$, $p < .005$] targets. There was no difference in detectability between same- versus different-location bimodal targets [$t(14) = 1.36$, *ns*]. In addition, right-sided targets were detected more accurately than left targets, both for visual [$t(14) = 5.66$, $p < .0001$] and auditory [$t(14) = 4.56$, $p < .001$] unimodal stimuli. False alarm rates were uniformly low: 1.5% for auditory, 1.0% for visual, 2.2% for bimodal standards with identical locations, and 2.0% for bimodal standards with different spatial locations. These false alarm rates did not differ significantly.

Reaction Time

Paralleling the detection accuracy results, reaction times (RTs) to visual targets were faster than RTs to auditory targets [$t(14) = 6.13$, $p < .001$], and RTs to the bimodal targets were markedly facilitated when compared to either auditory [$t(14) = 13.16$, $p < .0001$] or visual [$t(14) = 13.5$, $p < .0001$] targets. Overall, RTs did not differ significantly between same- and different-location bimodal targets [$t(14) = 0.91$, *ns*]. However, paired target stimuli originating from different locations were responded to faster when the visual stimulus was on the left [$t(14) = 3.84$, $p < .01$], and RTs to bimodal targets with identical locations were faster for right stimuli [$t(14) = 5.52$, $p < .001$]. The RTs for right unimodal visual targets were faster than for left visual targets [$t(14) = 3.73$, $p < .01$], but there were no differences due to stimulus laterality in the auditory modality [$t(14) = 0.94$, *ns*].

Table 1. Behavioral Measures of Hit Rate (% Correct) and RTs for Target Detections of Unimodal and Bimodal Stimuli

Targets	*Hit Rate (%)*	*SEM*	*RT (msec)*	*SEM*
V_L	60.0	2.6	502	7.2
V_R	77.5	1.1	473	8.1
A_L	46.3	2.5	549	9
A_R	57.5	2.6	541	9.3
A_LV_R	76.3	4.1	451	6.7
A_RV_L	81.1	1.2	435	6.8
A_LV_L	80.7	1.3	455	8.5
A_RV_R	79.9	4.2	423	5.9

Test of the Race Model

In order to verify that the speeding of RTs to bimodal stimuli resulted from a true facilitation of processing rather than an independent probability summation, the RT data were tested for violations of the independent race model (see Methods). The cumulative probability (CP) curves for RT revealed clear violations of the race model for bimodal targets at both the same and different locations (Figure 2, see Methods). In both cases, the observed bimodal CP_{AV} values for the faster RTs exceeded those predicted by the independent race model ($CP_A + CP_V - [CP_A \times CP_V]$). This difference between the observed and predicted values (the Miller ''inequality'') was significant over RT percentiles 20–40 for same-location targets ($2.2 < ts < 5.9$, $.0001 < ps < .05$) and over RT percentiles 10–40 for different-location targets ($2.2 < ts < 9.6$, $.0001 < ps < .05$). These Miller inequality functions did not differ significantly between the same-location and different-location targets.

ERP Results

The ERP components in the original waveforms elicited by auditory (A), visual (V), and bimodal (AV) standard stimuli are illustrated in Figure 3 for each of the four combinations of AV pairings. The auditory ERPs included typical P1 (peaking at around 60 msec), N1 (105 msec) and P2 (190 msec) components that were largest over central and frontal scalp areas. The visual ERPs included prominent P1 (120 msec), N1 (180 msec), and P2 (230 msec) components having maximal amplitudes over the posterior scalp. Both the auditory- and visual-evoked components could be discerned in the ERP to the bimodal stimuli.

Cross-modal interactions were evident in the AV − (A + V) difference waves, both when the bimodal AV stimuli were presented at the same (S) location and at different (D) locations (see Methods for details of how interactions were calculated). The same-location difference waves (Figure 4, left) showed a broad, posteriorly distributed positivity over the interval 160–220 msec (P190) and an anteriorly distributed negativity

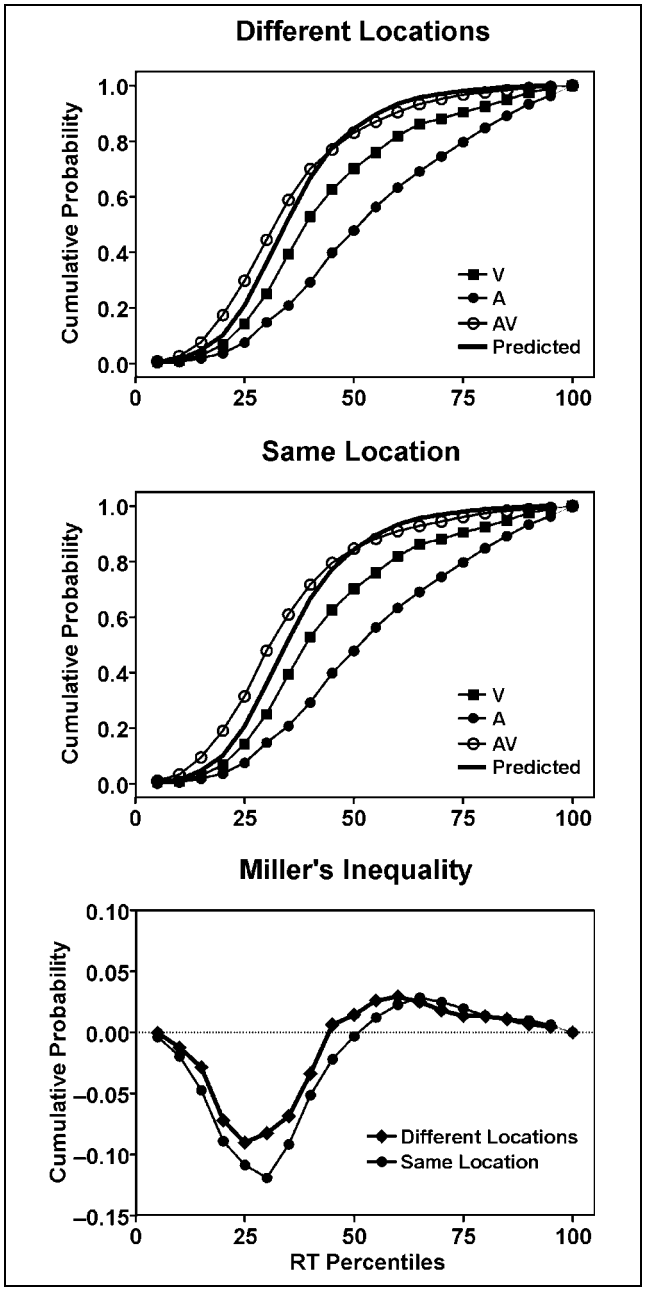

Figure 2. CP curves of RT distributions to unimodal visual (V), auditory (A), and bimodal (AV) targets for trials where bimodal targets occurred at the same and at different locations. Same-location distributions are averaged over A_LV_L and A_RV_R trials, and different-location distributions over A_LV_R and A_RV_L trials. Predicted curves show CP values predicted by the race model. Miller inequality curves plot the difference between observed bimodal CP values and the predicted values.

over 220–300 msec (N260) that were similar for the left-sided (S_L) and right-sided (S_R) difference waves. A combined analysis of the S_L and S_R difference waves showed that both the P190 and the N260 interaction components differed significantly from baseline: P190 [$F(1,14) = 10.7$, $p < .01$], N260 [$F(1,14) = 69.8$, $p < .001$]. Both the P190 and N260 tended to be larger over the hemisphere contralateral to the side of stimulation (see Figure 5), but this asymmetry only reached significance for the P190 [Condition S_L/S_R × Hemisphere interaction: $F(1,14) = 2.71$, $p < .05$].

The AV − (A + V) difference waves for stimuli at different locations (Figure 4, right) also showed prominent posterior P190 and anterior N260 components, along with an earlier N140 deflection. In a combined analysis of the D_{LR} and D_{RL} difference waves, all of these interaction components differed significantly from baseline: N140 [$F(1,14) = 4.81$, $p < .05$], P190 [$F(1,14) = 19.7$, $p < .001$], N260 [$F(1,14) = 65.3$, $p < .001$]. The N140 and P190 components tended to be larger over the hemisphere contralateral to the auditory stimulus (Figure 5), but these asymmetries did not reach significance.

Dipole modeling of the sources of these A–V interaction components was carried out separately for the same (S_L and S_R) and different (D_{LR} and D_{RL}) location conditions (Figure 6). The same-location P190 was accounted for by a pair of dipoles in the ventral occipito-temporal cortex, whereas the anterior N260 was well-fit by a superior temporal pair of dipoles. A very similar model provided a good account of the different-location interaction components; the N140–P190 sequence was accounted for by a ventral occipito-temporal dipole pair, and the N260 by a superior temporal pair of dipoles. The residual variance (RV) of these dipole models over the latency range 120–300 msec was 3.4% for the same-location and 3.2% for the different-location interaction components.

To test the significance of the apparent hemispheric asymmetries in the dipole source waveforms shown in Figure 6, *t* tests were carried out over successive 20-msec intervals over the range 100–400 msec comparing the S_L versus S_R and D_{LR} versus D_{RL} amplitudes for left and right hemisphere dipoles (see Methods). In the S_L versus S_R comparisons, the P190 was larger for the source contralateral to the side of stimulation over the intervals 140–180 msec (dipoles 1 and 2: ts > 2.62, ps < .01) and 180–200 msec (dipole 2: $t = 1.98$, $p < .03$); this asymmetry was reversed after 200 msec, however, with the P190 becoming larger at the ipsilateral source (dipoles 1 and 2: ts > 2.10, ps < .03). The N260 was also larger for the source contralateral to the side of stimulation over the interval 260–300 msec (dipoles 3 and 4: ts > 2.19, ps < .02). In the D_{LR} versus D_{RL} comparisons, the P190 was larger for the source contralateral to the auditory stimulus of the AV pair over the intervals 160–200 msec (dipoles 1 and 2: ts > 2.62, ps < .01); the N260 showed a similar pattern of asymmetry over the intervals 240–260 msec (dipoles 3 and 4: ts > 2.04, ps < .05) and 260–280 msec (dipole 4: $t = 2.16$, $p < .02$).

Differences in cross-modal interaction that depended on the spatial congruity of the auditory and visual stimuli were evident in the ''double-difference wave'' formed

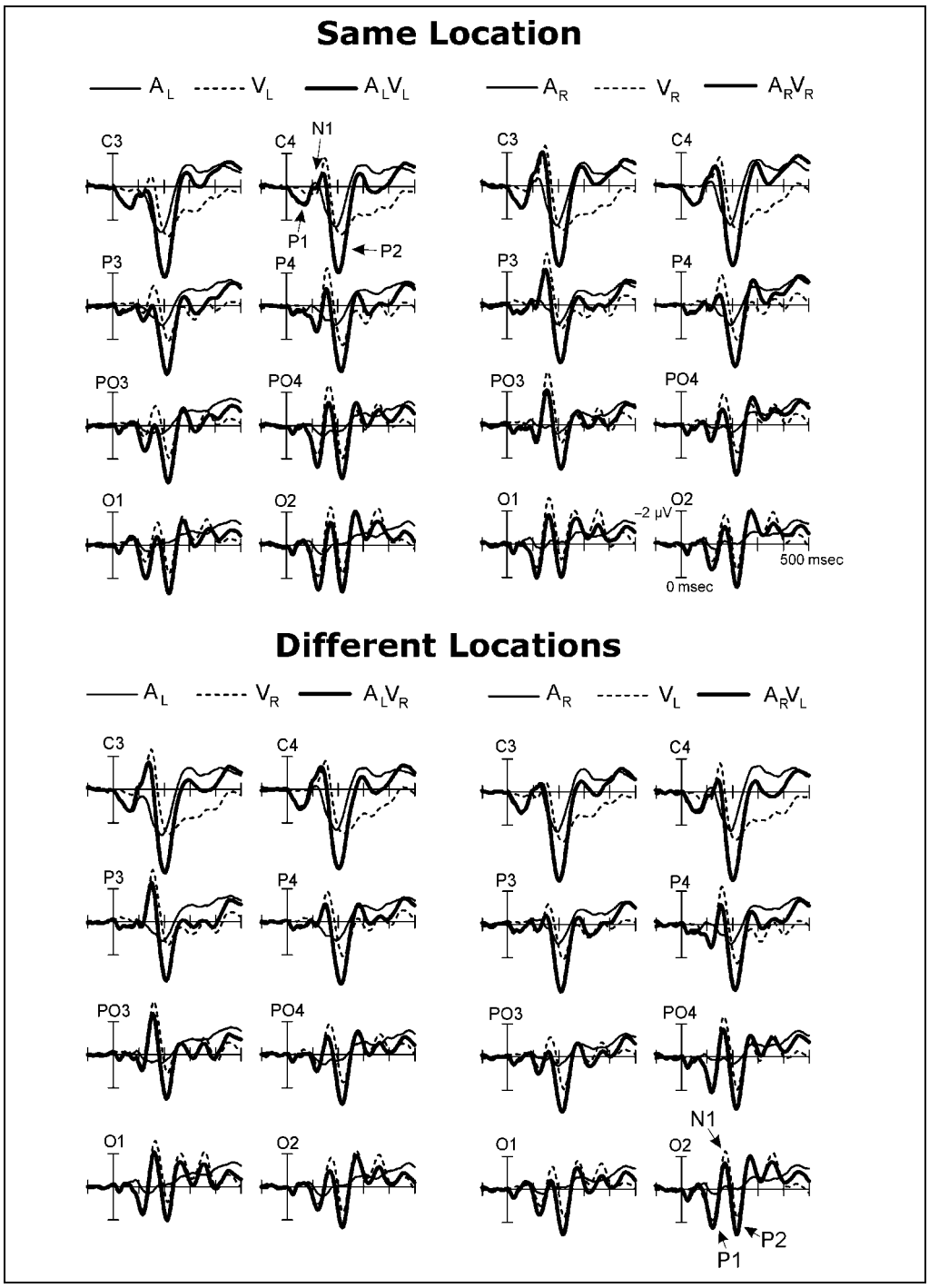

Figure 3. Grand-averaged ERPs to the different combinations of auditory (A), visual (V), and bimodal (AV) stimuli presented to the same or different left $_{(L)}$ and right $_{(R)}$ locations. Recordings shown are from eight selected scalp sites over occipital (O1, O2), parieto-occipital (PO3, PO4), parietal (P3, P4), and central (C3, C4) brain areas.

by subtracting the averaged same-location interaction waves ($S_L + S_R$) from the averaged different-location interaction waves ($D_{RL} + D_{LR}$). It should be noted that this double difference is algebraically equivalent to subtracting the summed ERPs to the same-location bimodal stimuli ($A_RV_R + A_LV_L$) from the summed ERPs to the different-location stimuli ($A_LV_R + A_RV_L$) (see Methods). The double-difference wave thus compares ERPs to the same physical stimuli when combined in bimodal pairs at different locations versus at the same location. As shown in Figure 7, the double-difference waveform (thick solid lines) includes significant posterior N140 [$F(1,14) = 8.01, p < .02$] and P190 [$F(1,14) = 6.55, p < .05$] components, resulting from these components being larger in the different-location ($D_{RL} + D_{LR}$) than in the same-location ($S_L + S_R$) difference waves. Following the N140 and P190 in the double-difference wave were a series of continuing posterior oscillatory deflections at approximately 10 Hz (N240, P290, N340, P390); of these, the P290 and P390 waves were significantly

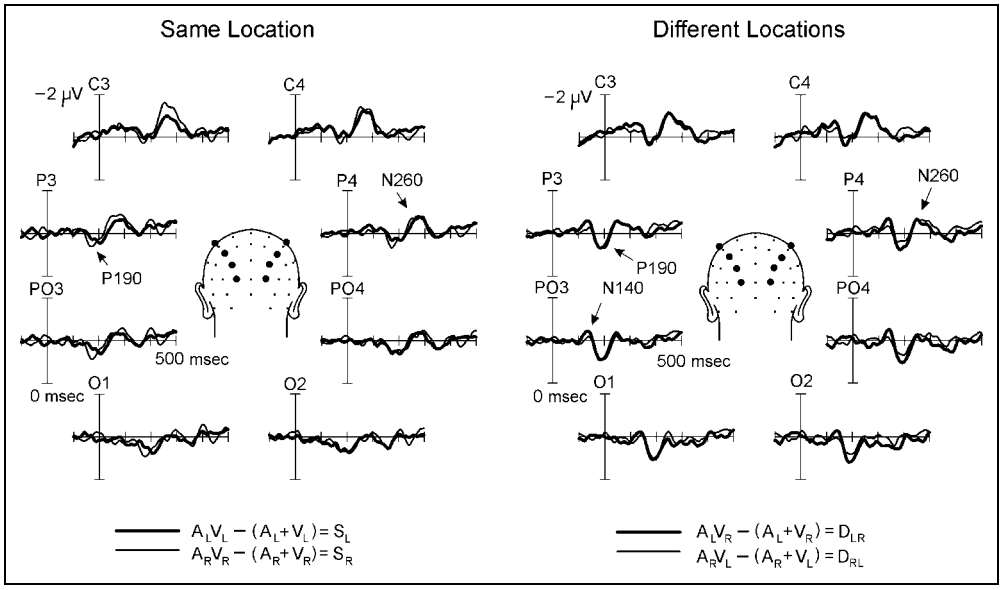

Figure 4. Grand-averaged difference waves formed by subtracting the sum of the unimodal auditory (A) and visual (V) ERPs from the bimodal (AV) ERP, for each of the different types of bimodal standard stimuli presented to the same (S) or different (D) locations.

different from baseline [$F(1,14) = 5.6$ and 5.1, respectively, both $p < .05$]. In contrast with the N140 and P190, the anterior N260 was of higher amplitude (in the interval 250–290 msec) in the same-location difference waves [$F(1,14) = 20.2$, $p < .001$].

To estimate the neural sources of these interaction components that were affected by spatial congruity, a dipole model was fit to the double-difference waveforms shown in Figure 7. The scalp distributions of the N140 and P190 of the posterior oscillatory sequence are shown in Figure 8 (top), along with that of the anterior P260 deflection resulting from the abovementioned difference in N260. A dipole pair situated in the ventral occipito-temporal region accounted for the N140–P190 and subsequent oscillations, whereas a dipole pair in the superior temporal cortex accounted for the P260 (Figure 8). The RV for this model over the interval 120–300 msec was 3.0%. Note the similarity of this dipole model accounting for the double-difference components with the models for the S_L/S_R and D_{RI}/D_{LR} interaction components (Figure 6).

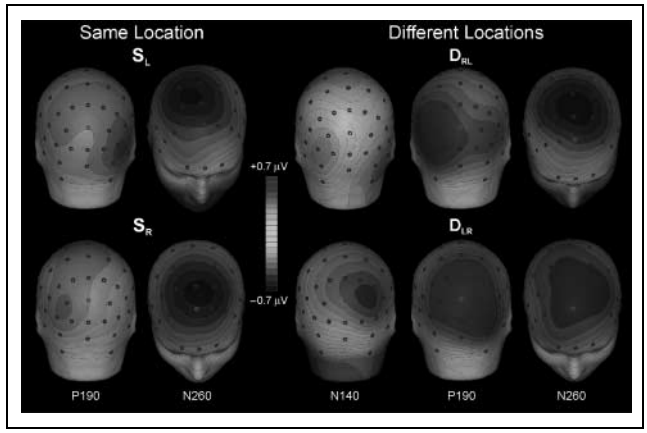

Figure 5. Topographical voltage maps of the AV − (A + V) interaction components seen in the difference waves of Figure 4. Note that for the N140 and N260, increasing darkness on the gray scale represents increased negativity.

To shed light on the genesis of the posterior oscillatory sequence in the double-difference ERP, its waveform can be compared with those of the summed same-location ($A_LV_L + A_RV_R$) and different-location ($A_LV_R + A_RV_L$) ERPs (Figure 9). It is evident that these summed bimodal ERPs have a very similar oscillatory waveform over the posterior scalp, which is attributable to their visual ERP constituent (note the similarity with the V_L and V_R waveforms in Figure 3). A small difference in the phase of these oscillations can be seen in these summed ERPs, with the phase of the $A_LV_R + A_RV_L$ waveform leading that of the $A_LV_L + A_RV_R$ waveform. This phase shift was evidenced by significantly shorter latencies in the peaks of the posterior oscillatory sequence in the summed different-location waveforms (Table 2). These slight latency (phase) differences do appear to be largely responsible for the oscillatory sequence N140–P190–N240–P290–N340–P390 that emerges when the $A_LV_L + A_RV_R$ waveform is subtracted from the $A_LV_R + A_RV_L$ waveform in the double-difference wave (Figure 9, thick solid lines). The N140 of this sequence could also be attributed in part to an enlarged positivity in the P1 latency range (100–150 msec) in the same-location summed waveform with respect to the different-location summed waveform.

The neural generators of these summed bimodal ERPs were estimated by dipole modeling (Figure 10). In both cases, the posterior oscillatory waves (P1, N1, P2, etc.) corresponding to the visual ERP could be accounted for by a ventral occipito-temporal dipole pair, whereas the auditory ERP components (N1, P2)

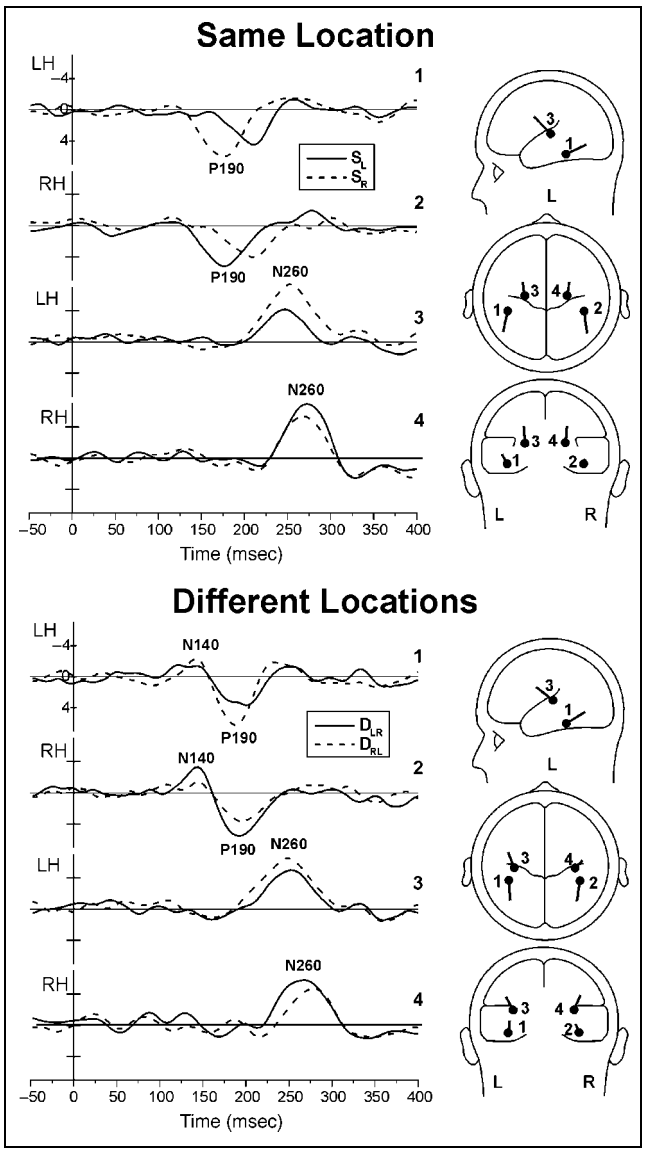

Figure 6. Dipole models of the AV − (A + V) difference waveforms shown in Figures 4 and 5. Dipole pair 1–2 was fit over the interval 130–210 msec, and dipole pair 3–4 was fit over 230–290 msec. Dipole source waveforms at the left show the interaction components that were modeled.

were well-fit by a superior temporal dipole pair. The RV for these dipole models over the interval 100–300 msec was 2.8% for the same-location summed ERP and 3.0% for the different-location summed ERP. Note the similarity of this model with those of the interaction components (Figure 6) and the double-difference components (Figure 8).

DISCUSSION

The main focus of this study was to find out whether multisensory integration of simultaneous AV stimuli would depend upon their spatial congruity. Behavioral evidence for such integration was obtained from target detection rates and RTs to targets presented either unimodally (A or V alone) or bimodally (AV pairings). Detection rates were found to be higher and RTs faster for the bimodal targets than for the corresponding unimodal targets for both spatially congruous and incongruous AV pairings. The degree of facilitation of RTs to bimodal AV targets was found to exceed simple probability summation as predicted by the independent race model (Miller, 1982). Given the numerous previous reports of greater behavioral and neural response facilitation for spatially congruous AV pairings (see Introduction), our finding that behavioral indices of AV integration were very similar for same- and different-location pairings was unexpected. Not only were there no significant differences in RTs or target detection rates for AV targets at same versus different locations, but the functions describing violations of the race model were nearly identical in the two cases.

Integrative interactions in the brain between the auditory and visual stimuli were defined in terms of the difference ERPs formed by subtracting the sum of the unimodal ERPs to A and V stimuli from the ERP to the simultaneous AV pair. Two major interaction components were found in the AV − (A + V) difference

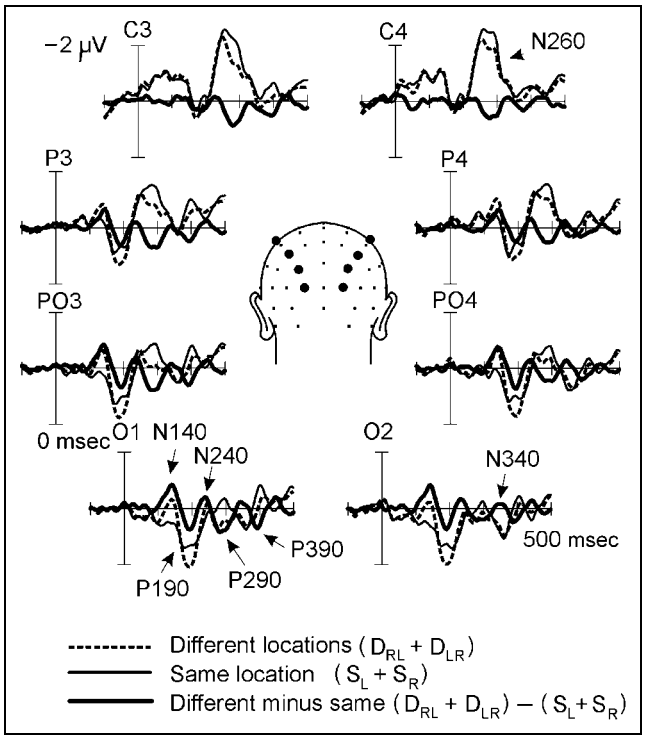

Figure 7. Effect of spatial congruity on cross-modal interaction as seen in the ''double-difference'' wave formed by subtracting the summed difference waves for same-location AV pairs (S_L + S_R) from the summed difference waves for different-location AV pairs (D_{RL} + D_{LR}).

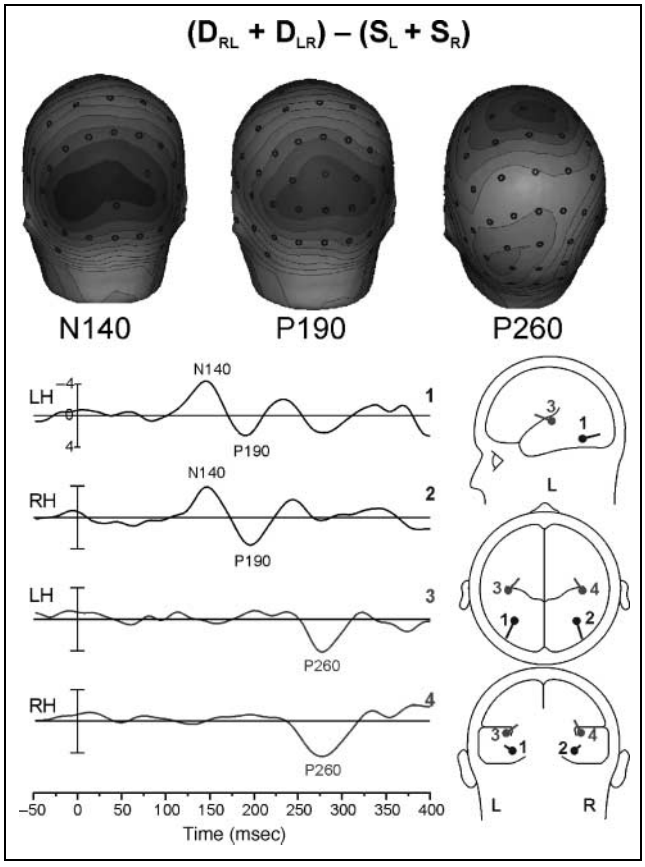

Figure 8. Dipole model of the neural sources of the components of the double-difference waveform $(D_{RL} + D_{LR}) - (S_L + S_R)$, shown as thick lines in Figure 7. Topographical voltage maps show the narrow occipital distributions of the initial N140 and P190 components of the posterior oscillatory sequence and the anterior distribution of the P260 deflection. Dipole pair 1–2 was fit over 130–200 msec and dipole pair 3–4 over 240–300 msec. Note: gray scale is same as in Figure 5.

waves regardless of the spatial congruity of the A and V stimuli. The first was a broad positive deflection in the interval 160–220 msec (P190) that was localized by dipole modeling to sources in the ventral occipito-temporal cortex, and the second was a broad anterior negativity at 220–300 msec (N260) that was localized to the superior temporal region. For AV pairs presented at the same location, the early part of the P190 (before 200 msec) was larger in amplitude in the hemisphere contralateral to the side of stimulation, whereas the later part (after 200 msec) was larger ipsilaterally; the N260 was consistently larger in amplitude in the hemisphere contralateral to the side of stimulation. In the case of different-location pairings, very similar P190 and N260 components were evident in the AV − (A + V) difference waves, with amplitudes that tended to be larger in the hemisphere contralateral to the auditory member of the AV pair. In addition, the different-location AV pairings elicited a posterior 10-Hz oscillatory ERP sequence that had earlier component latencies (i.e., was phase shifted) with respect to the same-location AV pairings. This phase-shifted oscillatory sequence was also localized by dipole modeling to sources in the ventral occipito-temporal cortex. In addition, the anterior N260 was found to be larger in amplitude for the same-location than for the different-location AV pairings.

In an experiment having a design similar to that of the present study, Gondan et al. (in press) observed an ERP interaction component in the $(A_LV_L + A_RV_R)$ minus $(A_LV_R + A_RV_L)$ difference waveform with a latency of 150–180 msec and a posterior scalp distribution. This effect was interpreted as reflecting a spatially dependent AV interaction in the parietal cortex, although a formal source localization analysis was not carried out. Based on its polarity and scalp topography, this interaction component most likely corresponds to the N140 of the present study, which initiated an oscillatory sequence of waves that we localized to the ventral occipito-temporal region. Gondan et al. did not examine the individual AV − (A + V) difference waves, so it is not clear whether common ERP interactions for spatially congruent and incongruent pairings were present. Their RT data, however, like those reported here, showed strong facilitative interactions for both same- and different-location pairings. Unlike the present study, Gondan et al. did find a small, but significant, speeding of RTs for the spatially congruent versus

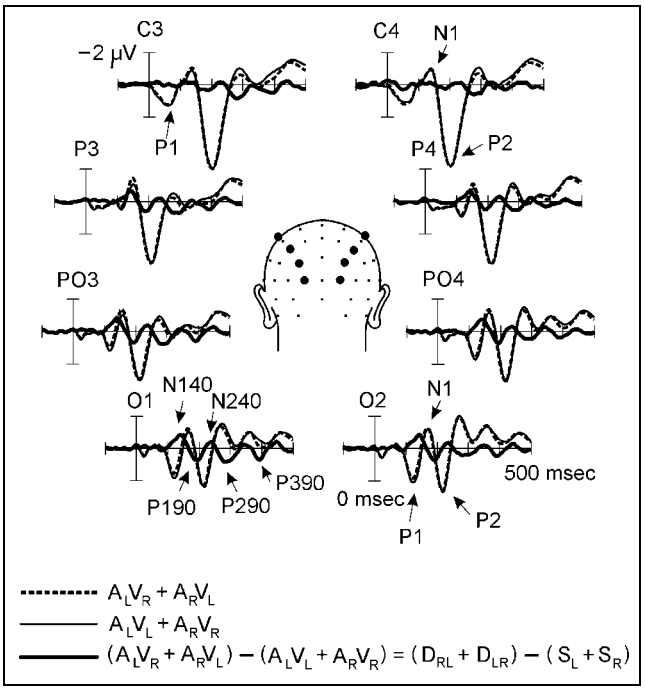

Figure 9. Comparison of the double-difference waveform with the summed bimodal ERPs $(A_LV_L + A_RV_R$ and $A_LV_R + A_RV_L)$. Labels of components at site O1 pertain to the double-difference wave, whereas labels at site O2 identify visual ERP components and labels at sites C3 and C4 identify auditory ERP components using same terminology as in Figure 3.

Table 2. Phase Shift Analysis

Peak Latencies (msec, SEM)		*Difference*	*Statistics*	
$A_LV_R + A_RV_L$	$A_LV_L + A_RV_R$	*(msec)*	*t(14)*	*p*
118 (1.8)	120 (1.8)	2	2.9	<.02
161 (2.2)	164 (2.4)	3	7.6	<.0001
212 (2.1)	215 (2.1)	3	3.6	.003
271 (2.0)	277 (2.4)	6	3.2	.006
316 (2.5)	321 (2.7)	4	2.7	.02
359 (2.6)	363 (2.7)	4	4.3	.0007
398 (3.8)	411 (2.7)	13	5.9	<.0001
459 (3.3)	464 (3.2)	5	5.3	.0003

incongruent AV pairings, which amounted to about a 1% difference.

The present study observed two ERP interaction components that were common to spatially congruent and incongruent AV pairings, namely, the occipito-temporal P190 and the superior temporal N260 waves. Previous studies have reported interaction components in the AV − (A + V) difference waveform for centrally presented stimuli that closely resembled the P190 (Fort et al., 2002; Molholm et al., 2002; Giard & Peronnet, 1999) and N260 (Teder-Sälejärvi et al., 2002) components observed here. The present behavioral results suggest that these two components reflect facilitatory AV interactions in visual (P190) and auditory or polymodal (N260) cortical areas, which occur regardless of spatial congruity. In a similar vein, Murray et al. (2005) found that behavioral and electrophysiological interactions between auditory (A) and somatosensory (S) stimuli were equivalent for spatially aligned and misaligned AS pairings. These results, together with those of Gondan et al. (in press) and the present study, suggest that the facilitation of behavioral responses produced by bimodal stimulation is mediated primarily by an energy summation mechanism (Welsh & Warren 1986) that has little or no dependence on the spatial congruity of the stimuli.

The P190 interaction component appears to represent an influence of simultaneous auditory input on visual processing in the ventral occipito-temporal pathways, which are generally considered specific to the visual modality. It has been proposed that such cross-modal influences are mediated by feedback influences from polymodal cortical areas, and earlier interaction components that may reflect such mediation have been reported in previous studies (McDonald, Teder-Sälejärvi, Di Russo, et al., 2003; Fort et al., 2002; Molholm et al., 2002; Giard & Peronnet, 1999). Although P190 interactions were present for all AV combinations, there were hemispheric asymmetries in their source waveforms that depended on stimulus location. In particular, the P190 was initially larger in the hemisphere contralateral to the same-location AV pairings and then shifted to the ipsilateral hemisphere. For the different-location pairings, the P190 was larger in the hemisphere contralateral to the auditory stimulus. This pattern of asymmetry suggests that the ventral visual cortex of each hemisphere receives a lateralized input from contralateral auditory space that initially affects visual processing to a greater extent in that hemisphere. An analogous asymmetry was reported by Murray et al. (2005), who found that AS interactions were localized to the auditory cortex contralateral to the stimulated hand for both same- and different-location pairings.

Previous studies have noted that the positive ERP interaction component corresponding to the P190 occurs at a latency similar to that of the major N1

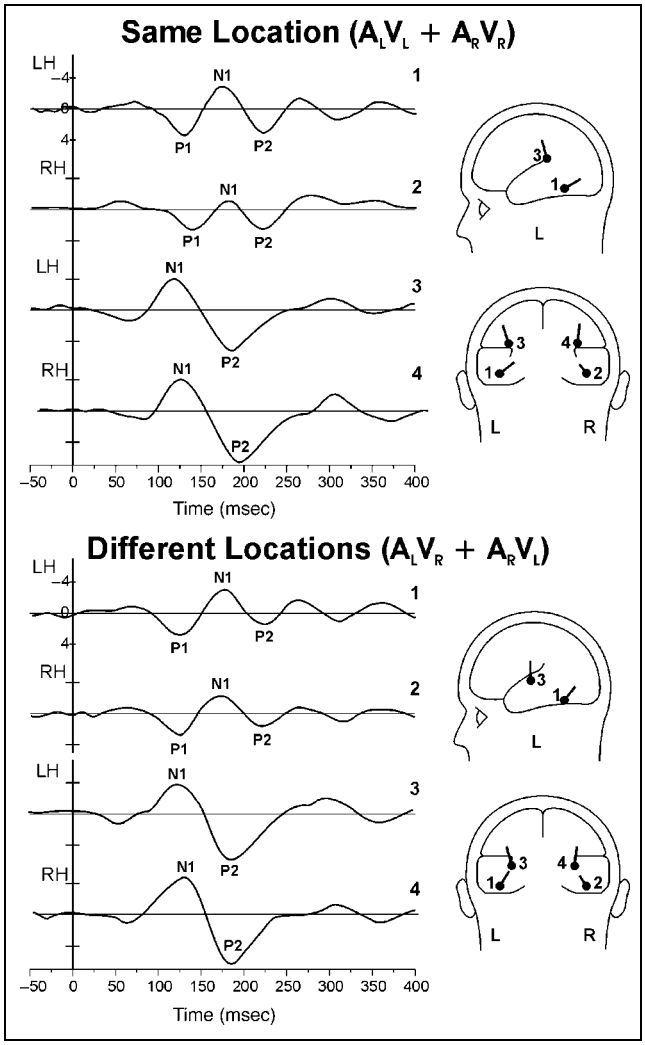

Figure 10. Dipole models of the summed bimodal ERPs with A and V stimuli presented at same and different locations. Dipole pair 1–2 was fit over the interval 120–150 msec, and dipole pair 3–4 was fit over 170–220 msec. Then both dipole pairs were refit over 120–300 msec.

component of the visual ERP and has a similar posterior scalp distribution (Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Giard & Peronnet, 1999). Accordingly, Giard and Peronnet suggested that this AV interaction was attributable to a reduction in amplitude of the visual-evoked N1, an example of cross-modal depression or subadditive interaction. An N1 reduction can also be seen in the present data (e.g., compare the AV and V waveforms in Figure 3), but the P190 positivity actually appears broader than the N1 and extends into the latency range of the visual-evoked P2. The functional significance of this cross-modal modulation of visual processing is unclear, but Giard and Peronnet suggested that the presence of a simultaneous auditory stimulus might reduce the need to engage attention to the visual stimulus, thereby reducing N1 amplitude. This accords with proposals that the amplitude of the visual N1 varies with the effectiveness of the stimulus in engaging attention (Arnott, Pratt, Shore, & Alain, 2001). A further suggestion along these lines would be that a spatially disparate sound might rapidly attract attention to its location and thereby further reduce the amplitude of the N1 to the visual stimulus in the opposite field. Such an effect might explain why the P190 effect (presumably resulting in part from a depression of the N1) was larger for different-location than for same-location AV pairings. It is debatable, however, whether attention could shift rapidly enough to one of two simultaneously presented stimuli so as to modulate processing of the other stimulus (Macaluso et al., 2000, 2001; McDonald, Teder-Sälejärvi, & Ward, 2001). In any case, the behavioral data showing cross-modal enhancement of performance suggest that this modulation of visual-evoked activity reflected in the P190 has a net facilitatory effect on discriminative signal processing.

In addition to the P190 and N260 interaction components, the spatially incongruous AV pairs produced a phase shift in a 10-Hz oscillatory sequence of visual ERP components localized to the ventral occipito-temporal cortex. The genesis of this oscillatory sequence was investigated by comparing its waveform and estimated dipolar sources with those of the summed bimodal ERPs to the same-location ($A_LV_L + A_RV_R$) and different-location ($A_LV_R + A_RV_L$) pairings. Both bimodal sums included sequences of visual-evoked components (P1, N1, P2, etc.) during the interval 100–400 msec that had a basic 10-Hz oscillatory waveform. The oscillatory sequence (N140, P190, etc.) seen in the difference wave formed by subtracting the same-location sum from the different-location sum appeared to result primarily from a slight phase difference in the 10-Hz visual-evoked components between these two summed waveforms. An enlarged positivity in the latency range of the visual P1 (100–150 msec) in the same-location sum also appeared to make a contribution to the N140. The hypothesis that the oscillatory sequence in the difference wave results from amplitude and phase modulations of the visual-evoked potential is supported by their having nearly identical ventral occipital sources (compare Figures 8 and 10).

Previous studies have identified several types of cross-modal interactions including facilitation of unimodal responses, depression of unimodal responses, and initiation of new activity specific to bimodal stimulation (e.g., Calvert et al., 2000; Raij et al., 2000; Giard & Perronet, 1999; Stein & Meredith, 1993). The phase shift of the visual ERP brought about by a spatially discrepant auditory stimulus appears to represent yet another type of cross-modal interaction. At present, we can only speculate on the functional consequences of this phase shift. One proposal would be that the phase shift is associated with a shift in the perceived location of the flash towards the location of the contralateral sound. This proposal is consistent with findings that the relative phase of visual ERP components between the two hemispheres shifts according to stimulus position in the lateral visual fields (Di Russo, Martinez, & Hilyard, 2003; Mangun, 1995). Behavioral studies have shown that a spatially discrepant auditory stimulus can indeed bias the perceived location of a visual stimulus toward its position, although not to the same extent as the visual bias on auditory localization (Bertelson & Radeau, 1981). Another possibility, not mutually exclusive with the foregoing, is that the phase shift could be a consequence of a rapid, automatic shift of attention to the auditory stimulus. McDonald, Teder-Sälejärvi, Di Russo, et al. (2003), found that oscillatory 10-Hz components in the visual ERP were shifted in phase as a function of the position of an attention-grabbing sound that preceded a flash by 120–300 msec. As noted above, however, it is not clear whether such an attentional shift could occur rapidly enough to influence visual ERP components to a simultaneous flash as in the present design.

Multisensory integration in the superior temporal cortex was evident in the form of a broad negativity (N260) in the AV − (A + V) difference wave that peaked over central scalp areas. A similar negative interaction component with a somewhat shorter latency (220–250 msec) was reported by Teder-Sälejärvi et al. (2002) in a task where stimuli were presented at a central location. The N260 was localized to the same general region of the superior temporal cortex as the auditory-evoked N1 component (compare Figures 6 and 10), and it appears to represent an influence of the simultaneous AV pairing on long-latency activity in either auditory-specific or polymodal areas of the superior temporal lobe. The N260 tended to be larger over the hemisphere contralateral to the auditory stimulus for both spatially congruent and incongruent AV pairings, in line with the contralateral predominance of the human auditory pathways. The finding that N260 was reduced in amplitude for different-location versus same-location AV pairings suggests a possible relationship to the well-

known biasing of perceived sound location by a concurrent visual stimulus (i.e., the ventriloquism effect). This is in line with current conceptions of cortical areas in the superior temporal plane being specialized for sound localization (Kaas, Hackett, & Tramo, 1999). This hypothesized role for the N260 as a sign of biased sound localization would imply that visual influences on sound localization occur at a fairly late stage of integrative processing.

Summary and Conclusions

The present results indicate that AV multisensory convergence is based upon some integrative neural processes that depend on the spatial congruity of the A and V stimuli and other processes that do not. The behavioral measures that were obtained showed strong cross-modal facilitation of target detectability and RT, but this facilitation of performance occurred regardless of the spatial coincidence of the A and V stimuli. Possible neural correlates of this spatially independent facilitation were identified as long-latency ERP components in ventral visual (P190) and superior temporal (N260) cortical areas that were elicited by both same-location and different-location AV pairings. These two AV interaction components appeared to arise from different mechanisms, however; whereas the P190 was a consequence of modulation of the visual-evoked potential (including a depression of the N1 component), the N260 appeared as new activity that did not have an obvious counterpart in the unimodal ERPs.

Despite the absence of a spatial congruity effect on behavioral indices of cross-modal integration, the neural response patterns to same-location and different-location AV pairings showed clear differences. The cross-modal interaction for different-location AV pairings included a phase shift of an oscillatory 10-Hz wave sequence elicited during the interval 100–400 msec when compared with the same-location pairings. This sequence appeared to represent an influence of spatially discrepant auditory input on early evoked activity in the ventral visual pathways that included a phase shift in the components of the visual ERP together with an amplitude modulation of its P1 component. An additional effect of spatial congruity was seen in the superior temporal N260 interaction component, which was larger in amplitude for same-location AV pairings. Considering that these ERP interactions were not associated with any differences in behavioral indices of target detectability or RT, it may be suggested that they reflect neural processes other than the cross-modal facilitation of signal strength. A reasonable inference, subject to future investigation, would be that these ERP interactions that depend on spatial congruity reflect neural activity associated with shifts in the perceived locations of the auditory and visual inputs that vary as a function of their spatial disparity.

METHODS

Participants

Fifteen healthy adults (eight women; ages 20–27 years, mean age 22.4 years) participated in this study after giving written informed consent, 14 of the subjects were right-handed. Each participant had normal or corrected-to-normal vision and was tested in the laboratory to confirm normal hearing.

Stimuli and Apparatus

The experiment was conducted in a sound-attenuated chamber with a background sound level of 32 dB (A) and a background luminance of 2 cd/m^2. Participants sat on a comfortable chair and faced two loudspeakers spaced 60° apart with red light-emitting diodes (LEDs) mounted on top of their cones (see Figure 1). Participants were instructed to maintain their gaze on a central fixation point throughout each experimental run and to press a button to infrequent target stimuli that could occur in either the auditory or visual modality (or both).

Randomized stimulus sequences consisting of A stimuli (500–15,000 Hz ''pink'' noise bursts, at 76 dB SPL), V stimuli (LED flashes, 75 cd/m^2), and AV stimuli (the simultaneous occurrence of noise burst and LED flash) were presented at irregular intervals of 300–600 msec. The eight possible stimulus combinations shown in Figure 10 occurred with equal probability in the sequence. All stimuli were of 33-msec duration, and infrequent target stimuli in both modalities had a slightly higher intensity than the frequent standard stimuli. The auditory target stimuli were noise bursts of increased loudness (by 3 to 9 dB), and the visual targets were flashes with increased brightness. The audiovisual target stimuli were always a combination of an auditory target and a visual target. Targets occurred infrequently ($p = .15$) for each of the eight stimulus configurations. In an initial practice run, the task difficulty (i.e., target discriminability) was individually adjusted to about 75% correct responses for both auditory and visual targets.

Each volunteer participated in one 1.5-hr session that consisted of an initial practice run and 16 experimental runs each consisting of 300 stimuli. Between runs, participants rested for 1–5 min. All procedures were approved by the UCSD Institutional Review Board for the protection of human subjects.

Behavioral Data Analysis

The percentage of correct target detections (hits) and mean RTs were calculated for each of the eight stimulus combinations for each individual subject. Additional tests were carried out to determine whether the antic-

ipated speeding of RTs to bimodal AV targets resulted from a true facilitation of processing due to bimodal interaction as opposed to the simple probability summation of target information arriving over two independent channels. According to the race model of bisensory processing, a speeding of RTs to bimodal targets can occur even though the unimodal inputs are processed independently (Miller, 1982). This follows from the assumption that the faster of the two inputs initiates the motor response on a given trial, and the likelihood of either of the two inputs producing a fast RT is greater than that of either input alone. Miller (1982) described a method for calculating whether the observed RTs to bimodal stimuli are facilitated over and above the predictions of the race model. As applied here, the procedure is to calculate the CP distributions of the RTs to unimodal A and V targets and bimodal AV targets. The independent race model specifies that for any given latency of RT, the bimodal CP value is less than or equal to the sum of the unimodal CP values minus their product, that is, $CP_{AV} < (CP_A + CP_V) - (CP_A \times CP_V)$. To the extent that the observed CP_{AV} values exceed those predicted by the race model, a mechanism of facilitative interaction between the unimodal inputs is supported. To make this comparison, CP values were determined over successive 5-percentile increments of the observed unimodal and bimodal RT distributions, and *t* tests were used to compare the observed CP_{AV} values with those predicted by the race model.

EEG Recording and Data Analysis

Electroencephalographic signals were recorded from 64 tin electrodes, including 56 sites from the 10-10 system (FPz, FP1, FP2, Fz, F1, F2, F3, F4, F5, F6, F7, F8, FCz, FC1, FC2, FC3, FC4, FC5, FC6, Cz, C1, C2, C3, C4, C5, C6, T7, T8, CPz, CP1, CP2, CP3, CP4, CP5, CP6, Pz, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, POz, PO3, PO4, PO7, PO8, Oz, O1, O2, Iz, and A1; American Electroencephalographic Society, 1994), and four additional electrodes located inferior to the occipital row of electrodes. Horizontal electrooculographic (EOG) signals were recorded bipolarly using electrodes at the left and right external canthi, and vertical EOG signals were recorded from an electrode below the left eye. All scalp electrodes, as well as the electrode below the left eye, were referenced to an electrode on the right mastoid (A2). Electrode impedances were kept below 5 kΩ.

All signals were amplified with a gain of 20,000 and a bandpass of 0.1–100 Hz (−12 dB/octave; 3 dB attenuation), digitized at a rate of 250 Hz, and stored on disk for off-line averaging. Automated artifact rejection was performed prior to averaging to discard trials during which an eye movement, blink, or amplifier blocking occurred. Signals from the remaining trials were averaged in 3000-msec epochs that started 1000 msec before each stimulus in order to allow for subsequent digital high-pass filtering. The averages were digitally low-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 46 Hz) to remove high-frequency noise produced by muscle movements and external electrical sources. In addition, the averaged waveforms were digitally high-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 2 Hz) in order to minimize the contribution of pre- and poststimulus slow potentials that may be confounded with early cross-modal interactions effects (Teder-Sälejärvi et al., 2002). The filtered averages were digitally re-referenced to the average of the mastoids.

Cross-modal interactions were investigated by subtracting the summed ERPs to the A and the V stimuli alone from the bimodal ERP to the combined AV stimuli, namely, the interaction = AV − (A + V). Only ERPs to the standard (nontarget) stimuli were included in these analyses. These difference waves were calculated for AV stimulus pairs sharing a common spatial location or coming from different spatial locations as follows:

Same-location difference waves:

Left: $A_LV_L - (A_L + V_L) = S_L$
Right: $A_RV_R - (A_R + V_R) = S_R$

Different-location difference waves:

Left: $A_LV_R - (A_L + V_R) = D_{LR}$
Right: $A_RV_L - (A_R + V_L) = D_{RL}$

A further comparison was made by subtracting the summed ''same-location'' difference waves from the summed ''different-location'' difference waves in order to reveal possible ERP signatures of spatial congruity. This ''double-difference'' subtraction is mathematically identical to the subtraction of the summed ERPs to the same-location bimodal stimuli ($A_RV_R + A_LV_L$) from the summed ERPs to the different-location stimuli ($A_LV_R + A_RV_L$). That is,

$$(D_{LR} + D_{RL}) - (S_L + S_R) = (A_LV_R + A_RV_L) - (A_LV_L + A_RV_R)$$

Several positive and negative peaks were identified in the AV − (A + V) difference waveforms between 100 and about 300 msec. For statistical testing, analyses of variance (ANOVAs) were carried out comparing mean amplitudes within specified time windows that included the peaks against the −100 to 0 msec prestimulus baseline. The following time windows were tested: 130–150 msec (N140), 170–210 msec (P190), and 230–290 msec (N260). For the N140 and P190 waves, amplitudes were measured at 16 sites in two clusters of eight electrodes per hemisphere. These parieto-occipital sites were for the left hemisphere P1, P3, P5, P7, PO3, PO7, O1, and I3, and for the right hemisphere P2, P4, P6, P8, PO4, PO8, O2, and I4, respectively. For measuring the

N260 wave, 16 electrodes were selected from frontocentral to centro-parietal sites. For the left hemisphere, these sites were F1, F3, FC1, FC3, C1, C3, CO1, and CP3. For the right hemisphere, the locations were F2, F4, FC2, FC4, C2, C4, CP2, and CP4. ERP amplitudes within these clusters were subjected to repeated-measures ANOVAs with the factors of stimulus condition (laterality) and hemisphere.

Source Localization

Estimation of the dipolar sources of ERP components and interactions was carried out using Brain Electrical Source Analysis (BESA 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 ERP distribution (Scherg, 1990). The algorithm interactively adjusts (fits) the location and orientation of the dipole sources in order to minimize the 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. The mean spherical coordinates for each site averaged across subjects were used for topographic mapping and source localization with respect to the standardized finite element model of BESA 2000. Symmetrical pairs of dipoles were fit sequentially to the distinctive components in the grand-average waveforms within specified intervals. Dipole pairs were constrained to be mirror-image dipole in location only. The dipole fitting strategy was as follows: For the same-location conditions, a dipole pair was fit to the P190 component and another pair to the N260 component. For the different-location conditions, a dipole pair was fit to the N140–N190 sequence of components and another pair to the N260 component.

To test whether lateral asymmetries in the grand-average dipole source waveforms (Figure 6) were significant, *t* tests were carried out over successive 20-msec intervals, comparing amplitudes of the source waveforms for S_L versus S_R and D_{LR} versus D_{RL} for left and right hemisphere dipoles. As estimates of noise variability, amplitude differences for the above comparisons were calculated for 25 time points in the grand-average source waveforms within a 100-msec prestimulus baseline period (Di Russo et al., 2003). Each *t* test ($df = 28$) compares the mean amplitude of the 25 baseline points with the mean of 5 points within each 20-msec measurement interval of the grand-average source waveforms.

Acknowledgments

This study was supported by a grant from the National Institute for Mental Health (MH 25594). We thank Daniel A.-J. Heraldez for technical assistance.

Reprint requests should be sent to Steven A. Hillyard, University of California, San Diego, Department of Neurosciences 0608, School of Medicine, La Jolla, CA 92093–0608, USA, or via e-mail: wat@sdepl.ucsd.edu.

REFERENCES

American Electroencephalographic Society. (1994). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 11,* 111–113.

Arnott, S. R., Pratt, J., Shore, D. I., & Alain, C. (2001). Attentional set modulates visual areas: An event-related potential study of attentional capture. *Cognitive Brain Research, 12,* 383–395.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41,* 809–823.

Bernstein, I. H., & Edelsein, B. A. (1973). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology, 87,* 241–247.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory–visual spatial discordance. *Perception & Psychophysics, 29,* 578–584.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10,* 649–657.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect. *Neuroimage, 14,* 427–438.

Corneil, B. D., & Munoz, D. P. (1996). The influence of auditory and visual distractors on human orienting gaze shifts. *Journal of Neuroscience, 16,* 8193–8207.

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

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Early auditory–visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research, 14,* 20–30.

Frens, M. A., Van Opstal, O. A., & Van der Willigen, W. R. (1995). Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements. *Perception & Psychophysics, 57,* 802–816.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research, 10,* 77–83.

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., Ritter, W., & Murray, M. M. (2002). Auditory–somatosensory multisensory processing in auditory association cortex: An fMRI study. *Journal of Neurophysiology, 88,* 540–543.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147,* 332–343.

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

Gondan, M., Niederhaus, B., Röder, B., & Rösler, F. (in press). Multisensory processing in the redundant target effect: A behavioral and ERP study. *Perception & Psychophysics.*

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual–auditory interactions in human sensorimotor processing. *Experimental Brain Research, 122,* 247–252.
Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual–auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 131–153.
Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in the primate cerebral cortex. *Current Opinion in Neurobiology, 9,* 164–170.
Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience, 14,* 420–429.
Loveless, N. E., Brebner, J., & Hamilton, P. (1970). Bisensory presentation of information. *Psychological Bulletin, 73,* 161–199.
Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science, 289,* 1206–1208.
Macaluso, E., Frith, C. D., & Driver, J. (2001). Reply to McDonald, Teder-Sälejärvi, & Ward. *Science, 292,* 1791.
Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology, 32,* 4–18.
McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by crossmodal spatial attention. *Journal of Cognitive Neuroscience, 15,* 10–19.
McDonald, J. J., Teder-Sälejärvi, W. A., & Ward, L. M. (2001). Multisensory integration and cross-modal attention effects in the human brain. *Science, 292,* 1791.
Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279.
Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research, 14,* 115–128.
Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2005). Grabbing your ear: Auditory–somatosensory multisensory interactions in early sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex, 15,* 963–974.
Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron, 28,* 617–625.
Scherg, M. (1990). Fundamentals of dipole source potential analysis. In: F. Grandori, M. Hoke, & G. L. Romani, (Eds.), *Auditory evoked magnetic fields and electric potentials. Advances in audiology* (Vol. 5, pp. 40–69). Basel: Karger.
Schröger, E., & Widman, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology, 35,* 755–759.
Schürmann, M., Kolev, V., Menzel, K., & Yordanova, J. (2002). Spatial coincidence modulates interaction between visual and somatosensory evoked potentials. *NeuroReport, 13,* 779–783.
Simon, J. R., & Craft, J. L. (1969). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of Experimental Psychology, 86,* 272–274.
Stein, B. E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research, 123,* 124–135.
Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 1,* 12–24.
Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses* (pp. 123–156). Cambridge: MIT Press.
Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. W. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience, 1,* 12–24.
Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event–related potentials. *Cognitive Brain Research, 14,* 106–114.
Welsh, R. B., & Warren D. H. (1986). Intersensory interactions. In: K. R. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance, Volume 1: Sensory processes and perception* (pp. 1–36). New York: Wiley.