

The Cross-Modal Spread of Attention Reveals Differential Constraints for the Temporal and Spatial Linking of Visual and Auditory Stimulus Events

Sarah E. Donohue,^{1,3} Kenneth C. Roberts,¹ Tineke Grent-‘t-Jong,^{1,5} and Marty G. Woldorff^{1,2,3,4}

¹Center for Cognitive Neuroscience and ²Department of Psychology and Neuroscience, Duke University, Durham, North Carolina 27708, Departments of ³Neurobiology and ⁴Psychiatry, Duke University, Durham, North Carolina 27710, and ⁵Department of Experimental Psychology and Psychopharmacology, University of Utrecht, Utrecht, 3584CS, The Netherlands

The integration of multisensory information has been shown to be guided by spatial and temporal proximity, as well as to be influenced by attention. Here we used neural measures of the multisensory spread of attention to investigate the spatial and temporal linking of synchronous versus near-synchronous auditory and visual events. Human participants attended selectively to one of two lateralized visual-stimulus streams while task-irrelevant tones were presented centrally. Electrophysiological measures of brain activity showed that tones occurring simultaneously or delayed by 100 ms were temporally linked to an attended visual stimulus, as reflected by robust cross-modal spreading-of-attention activity, but not when delayed by 300 ms. The neural data also indicated a ventriloquist-like spatial linking of the auditory to the attended visual stimuli, but only when occurring simultaneously. These neurophysiological results thus provide unique insight into the temporal and spatial principles of multisensory feature integration and the fundamental role attention plays in such integration.

Introduction

Sensory perception entails the ability to combine information from multiple modalities to reliably determine the objects in one’s environment. This can be accomplished in various ways, often invoking selective attention toward one or multiple modalities to guide the stimulus processing. Fundamental to the grouping of multisensory information into meaningful entities are several general principles that underscore the temporal and spatial linking of stimulus input components (Stein and Meredith, 1993; Stein and Stanford, 2008).

The necessity for some degree of temporal correspondence for multisensory integration to occur has been shown both neurophysiologically and behaviorally. More specifically, as the stimulus onset asynchrony (SOA) between the unisensory components of a multisensory stimulus increases (beyond the typical temporal window of integration of ~150 ms), the probability that they will be neurally integrated and judged as from the same source or event decreases (Meredith et al., 1987; Stone et al., 2001; Schneider and Bavelier, 2003).

Likewise, with both neural recordings (Meredith and Stein, 1986; Wallace et al., 1996) and behavioral measures (Spence et al., 2003; Gondon et al., 2005; Keetels and Vroomen, 2005; Bolognini et al., 2007), as spatial discrepancy increases, the likelihood also decreases for physiological multisensory interaction and for the behavioral judgment of perceptual correspondence. Importantly, however, as manifested in the phenomenon of ventriloquism, when a physically separated sound occurs concurrently with a visual stimulus, the perceived auditory location tends to be shifted toward the visual (Bertelson and Radeau, 1981; Hairston et al., 2003), and spatially separated but synchronous multisensory stimuli can still yield behavioral and neural enhancements (Teder-Sälejärvi et al., 2005). Moreover, in an explicit auditory localization task, this perceptual shift has been associated with a shift in the auditory brain response toward the side contralateral to the visual stimulus, presumably reflecting perceptual spatial integration of these spatially disparate stimuli (Bonath et al., 2007). To date, however, relatively little is known about how the temporal factors of multisensory integration interact neurally with the spatial factors, such as during the ventriloquist illusion, and the role attention might play in these interactions.

Previously, we had reported neural activity measures showing that attention to stimuli in one modality (vision) can spread to irrelevant but synchronous stimuli in another modality (audition), even when arising from different locations (Busse et al., 2005). This effect was reflected electrophysiologically by a late-onset (>200 ms), long-lasting, negative-polarity ERP wave, as well as enhanced auditory-cortex fMRI activity, being elicited by sounds occurring synchronously with an attended, spatially disparate, visual stimulus. This spreading-of-attention effect was

Received Oct. 8, 2010; revised Feb. 20, 2011; accepted April 1, 2011.

Author contributions: S.E.D. and M.G.W. designed research; S.E.D. performed research; K.C.R. contributed unpublished reagents/analytic tools; S.E.D., T.G.-‘t.J., and M.G.W. analyzed data; S.E.D., K.C.R., T.G.-‘t.J., and M.G.W. wrote the paper.

This work was supported by a National Science Foundation graduate research fellowship to S.E.D. and by an NINDS grant (R01-NS051048) to M.G.W. We thank Maria A. Pavlova for assistance with data collection and analysis.

Correspondence should be addressed to Dr. Marty G. Woldorff, Professor (Psychiatry; Psychology & Neuroscience; Neurobiology), Center for Cognitive Neuroscience, Duke University, Box 90999, Durham, NC 27708-0999. E-mail: woldorff@duke.edu.

DOI:10.1523/JNEUROSCI.5298-10.2011

Copyright © 2011 the authors 0270-6474/11/317982-09\$15.00/0

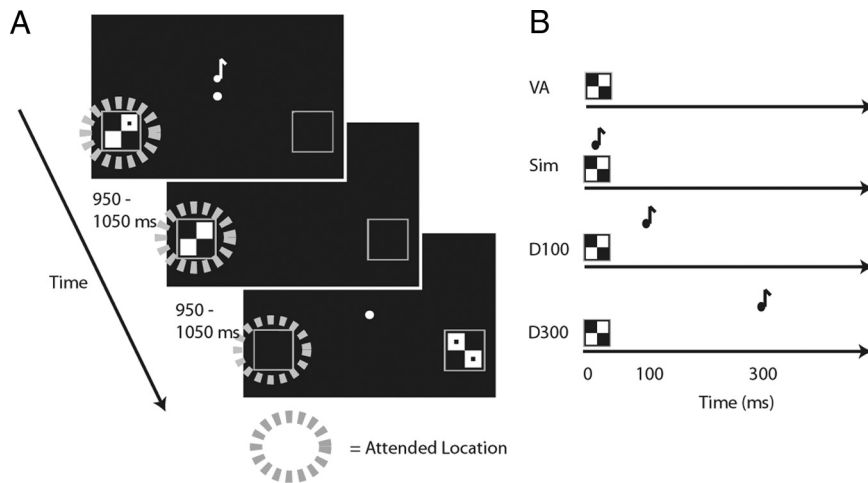


Figure 1. Task. **A**, Task timing and example stimuli. In a given run, while EEG data were collected, subjects were instructed to fixate centrally and to covertly visually attend to either the left or the right to detect infrequent targets (checkerboards with two dots) on that side. The lateral visual stimuli could be accompanied by a centrally presented tone at the varying SOAs (shown in **B**). The intertrial intervals were jittered between 950 and 1050 ms. **B**, Relative timing of the visual and auditory stimulus components in the 4 conditions: visual only, visual stimulus presented alone; simultaneous, visual and auditory stimuli presented simultaneously; delay-100, auditory stimulus delayed by 100 ms relative to the onset of the visual; and delay-300, auditory stimulus delayed by 300 ms relative to the visual. An additional post-EEG behavioral simultaneous-judgment task was conducted using the same stimuli as during all the multisensory EEG runs, but in which subjects were asked to judge whether the auditory and visual stimuli occurred simultaneously or separately.

interpreted as a cross-sensory, object-related, linking process (see also Molholm et al., 2007; Fiebelkorn et al., 2010), reflecting a multisensory version of attentional spreading observed across unimodal visual objects (Egley et al., 1994; Martinez et al., 2007). Here, we investigated the temporal and spatial linking of multisensory stimulus components, and the role of attention in this linking, by examining the cross-modal attentional spreading between spatially disparate visual and auditory stimulus events occurring with different temporal separations, both within and outside of the temporal window of integration. We report a fundamental role of attention in the multisensory-linking processes, as well as a dissociation between patterns of brain activity reflecting the temporal and spatial linking of the stimuli.

Materials and Methods

Participants. Eighteen healthy right-handed adult volunteers (9 male) participated in the study [ages, 18–24 years; mean (M) = 21.1]. Two additional participants were excluded because of poor behavioral task performance. Participants gave written informed consent and were financially compensated for their time. All procedures were approved by the Duke University Health System Institutional Review Board.

Stimuli and task. To determine the spread of attention at varying SOAs, we adapted the bilateral attentional streaming paradigm we had used previously for simultaneous visual and auditory events (Busse et al., 2005). During each block, participants were instructed to covertly visually attend to the left side or to the right side of a central fixation point (Fig. 1). Visual stimuli were randomly presented to the lower left or lower right quadrant of the screen (at 12.3° visual angle to the left and right of the center, and 3.4° below the central fixation). The visual stimuli were checkerboard images with 0, 1, or 2 dots contained within the checkerboard. Each visual stimulus was on the screen for 33 ms, and the intertrial interval was jittered between 950 and 1050 ms. Participants were instructed to detect an occasional target visual image (a checkerboard on the designated side with 2 dots, 14% probability) in the attended visual stream, and to press a button when this image appeared. Accuracy and reaction times (RTs) were recorded, and for each participant the difficulty level was titrated by adjusting the contrast and size of the dots

within the target images so that participants were ~80% correct in detecting the target-stimulus checkerboard possessing two dots.

All visual stimuli (targets and non-targets, both on the attended and unattended sides) were presented in the following multisensory conditions: visual only, visual with simultaneous auditory (simultaneous), visual with auditory delayed by 100 ms (delay-100), or visual with auditory delayed by 300 ms (delay-300; Fig. 1B). In each of the multisensory trial types, the auditory stimulus consisted of a tone pip (33 ms duration, 1200 Hz, 60 dB SPL, 5 ms rise and fall periods) presented centrally. Participants were instructed to ignore all of the auditory stimuli as they were irrelevant to their task. All stimuli were presented in Matlab (MathWorks) using Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). After one practice block, participants completed a total of 30 experimental blocks (half attend left, half attend right), each a little more than 2 min in duration. The trial types were presented in randomized and counterbalanced order within each block, and the order of the blocks was randomized for all participants.

Post-EEG behavioral assessment of simultaneous judgment perception. To assess participants' ability to determine the temporal separation between auditory and visual events, each participant was behaviorally tested using a simultaneity judgment task immediately after the EEG recording session.

In this task, as before, participants were instructed to covertly attend to the left or right side of a central fixation point, and lateralized visual streams were presented as during the EEG session. The auditory stimuli, also similar to before, were presented centrally, either simultaneously with the visual (simultaneous), delayed by 100 ms (delayed-100), or delayed by 300 ms (delayed-300), and the visual stimuli were always accompanied by one of these auditory conditions (i.e., there were no visual-only trials here). The time between consecutive visual stimuli was jittered from 1450 to 1550 ms, to allow enough time for participants to make a simultaneity judgment and to respond. More specifically, participants were instructed to judge whether the visual and auditory components of the stimulus were simultaneous, indicating their judgment with a button press. A total of 48 trials were completed for each of the 3 conditions.

Behavioral data analysis. RTs, hits, and false alarms were obtained for each participant for the behavioral data obtained during the EEG recording session. Outlier trials, as defined by having reaction times more than 2 SDs from each participant's mean reaction time, were excluded from the analysis. Repeated-measures ANOVAs were conducted to look at the effect of multisensory-SOA condition on reaction time and accuracy, and any significant effects, using an α level of 0.05, were followed up with *t* tests. In the simultaneity judgment task, the "percent simultaneous" judgment responses were calculated for each condition, and a repeated-measures ANOVA was conducted to see whether these judgments differed between the SOA conditions.

EEG recording and analysis. Continuous electroencephalogram (EEG) was recorded using a NeuroScan SynAmps system, and a customized elastic electrode cap (Electro-Cap International) contained 64 electrodes. The data were sampled at 500 Hz, bandpass filtered online from 0.01 to 100 Hz, and referenced to the right mastoid electrode site. Eye blinks and eye movements were recorded using two electrodes lateral to each eye, referenced to each other, and two electrodes inferior to each orbit, referenced to electrodes above the eyes.

Offline, the data were filtered with a low-pass filter that strongly attenuated signal frequencies >50 Hz. Trials that contained eye movements or blinks were rejected, as were trials with excess muscle activity or excess slow drift. The time range around each trial used for assessment of artifact was -250 to 950 ms, relative to the onset of the visual stimulus. The

artifact-rejection threshold level was titrated individually for each participant, and that value was then used for computer-applied rejection processes applied in an automatic fashion across all the trials for that participant. The data were re-referenced to the algebraic average of the left and right mastoid electrodes. Time-locked ERP averages were obtained for each of the different conditions, and difference waves were calculated based on these averages. For the analyses reported here, only the non-target trials were considered, thereby focusing on the influence of the visual spatial attention manipulation without the presence of the large, long-latency ERP waves (e.g., P300s) associated with target detection. To examine the differences between conditions, repeated-measures ANOVAs were conducted on mean-amplitude measures of ERP brain activity (see Results) across participants using a prestimulus baseline of 200 ms. All offline processing was done using the ERPSS software package (University of California, San Diego, San Diego, CA).

Extraction of spreading-of-attention activity. To extract the activity associated with the processing of the task-irrelevant auditory tone as a function of whether it was accompanied by an attended or unattended visual stimulus, the following analysis was conducted. In each of the three multisensory conditions (simultaneous, delay-100, and delay-300), the task-irrelevant tones were always presented centrally in the same trial with either a visually attended or a visually unattended lateral stimulus. To separate the contribution of the visual stimuli on the ERPs in the multisensory conditions, the visual-only condition was subtracted from the simultaneous, delay-100, and delay-300 conditions (all time-locked to the onset of the visual stimulus), separately for each visual-attention condition, isolating the activity linked to the processing of the auditory stimulus under each multisensory attentional context. These extracted ERP responses to the central tones when they were accompanied by an attended versus an unattended lateral visual event could then be compared to extract the possible spread of attention across modality and space to the tones (see Fig. 4). In addition, the conditions were collapsed across the left and right side to obtain this overall attentional spreading effect, regardless of the side of visual stimulation. This spreading-of-attention activity was extracted and analyzed for each of the three SOA delay conditions.

Two additional comparisons were performed between the extracted auditory responses for the three SOA conditions: one for when the lateral visual stimulus was attended and one for when it was unattended. These were to assess the effect of the SOA manipulation separately within each of the visual-attention conditions.

Results

Behavioral results for the visual attention task during the EEG runs

For the visual attention task during which the EEG was recorded, RTs and detection accuracy for the visual target stimuli were collected. (Note that the centrally presented auditory tones were always task irrelevant in these runs.) No significant differences in accuracy for the visual targets were observed between the three SOA conditions, with the performance in each case being close to the desired difficulty titration level of 80% correct (visual only: $M = 78.3\%$, $SD = 13.3\%$; simultaneous: $M = 78.8\%$, $SD = 9.5\%$; delay-100: $M = 77.0\%$, $SD = 14.1\%$; delay-300: $M = 78.8\%$, $SD = 12.5\%$). For the RTs (visual only: $M = 587$ ms, $SD = 46$ ms; simultaneous: $M = 583$ ms, $SD = 49$ ms; delay-100: $M = 586$ ms, $SD = 48$ ms; delay-300: $M = 595$ ms, $SD = 50$ ms), however, an ANOVA revealed a main effect of condition ($F_{(3,51)} = 2.82$; $p < 0.05$), with the delay-300 condition being significantly slower than the other conditions with an auditory component (simultaneous vs delay-300: $t_{(17)} = 2.53$; $p < 0.05$; delay-100 vs delay-300: $t_{(17)} = 2.90$; $p = 0.01$; visual only vs delay-300: $t_{(17)} = 1.96$; $p = 0.07$).

Simultaneity-judgment task (behavior only)

In the separate behavioral task after the EEG session, participants were highly likely to judge the visual and auditory events as si-

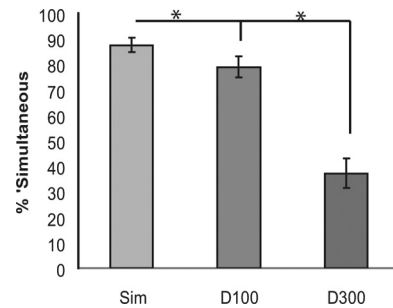


Figure 2. Behavioral results from the post-EEG behavioral simultaneity-judgment task. Shown are the simultaneous (Sim), tone-100-ms-delayed (D100), and tone-300-ms-delayed (D300) conditions (error bars show the SEM). When the stimuli were simultaneous or delayed by 100 ms, subjects were much more likely to judge the stimuli as occurring simultaneously than when they were presented 300 ms apart.

multaneous in both the simultaneous and delay-100 conditions (87.55% and 78.85%, respectively), while they were very unlikely to judge them as simultaneous in the delay-300 condition (37.16%; Fig. 2). An ANOVA revealed a main effect of condition ($F_{(2,34)} = 60.81$; $p < 0.001$), with *post hoc* *t* tests showing differences between all three conditions (simultaneous vs delay-100: $t_{(17)} = 4.11$; $p = 0.001$; simultaneous vs delay-300: $t_{(17)} = 8.68$; $p < 0.001$; delay-100 vs delay-300: $t_{(17)} = 7.21$; $p < 0.001$).

Visual spatial attention ERP effects

Visual spatial attention effects on the non-target visual stimuli that occurred by themselves (visual-only trials) were assessed to assure that the manipulation of the subjects' covert visual spatial attention was effective. Classical attentional modulations (Hillyard and Anllo-Vento, 1998) of the early sensory ERP components contralateral to the direction of visual attention were observed for both directed loci of attention (left and right). In particular, attended compared with unattended visual stimuli showed an increased positivity-polarity component at contralateral occipital sites (P1 effect) between 90 and 120 ms poststimulus, followed by an increased negativity-polarity wave over contralateral parietal-occipital sites (posterior N1 effect) between 190 and 230 ms (Fig. 3). An ANOVA that included the factors of attention (attended vs unattended), stimulus location (left vs right visual field), and hemisphere (left vs right electrode location) confirmed the presence of a significant contralateral P1 attention effect over the latency window 90–120 ms, with a three-way significant interaction across the occipital sites TO1/TO2, O1i/O2i, and P3i/P4i ($F_{(1,17)} = 22.87$ $p < 0.0005$). (Electrodes are labeled according to the International 10-20 system. For electrode locations that are close, but not identical to the standard 10-20 system locations, the postscript "i" and "a" are used to indicate a location slightly inferior or anterior to the standard location.) The analyses similarly showed a significant N1 attention effect at posterior sites P3i/P4i, P3a/P4a, and O1/O2 ($F_{(1,17)} = 5.02$ $p < 0.05$) from 190 to 230 ms. These attention effects on the sensory evoked ERP waves indicate that subjects were appropriately focusing their visual attention to the instructed side.

The cross-modal spread of attention from vision to audition as a function of SOA

Simultaneous condition

Looking first at the simultaneous condition, the ERPs to attended visual stimuli occurring alone were subtracted from the ERPs to attended visual stimuli occurring with a task-irrelevant central tone, thereby deriving an extracted ERP to the central auditory

Visual Attention Effects (Non-targets, VO Condition)

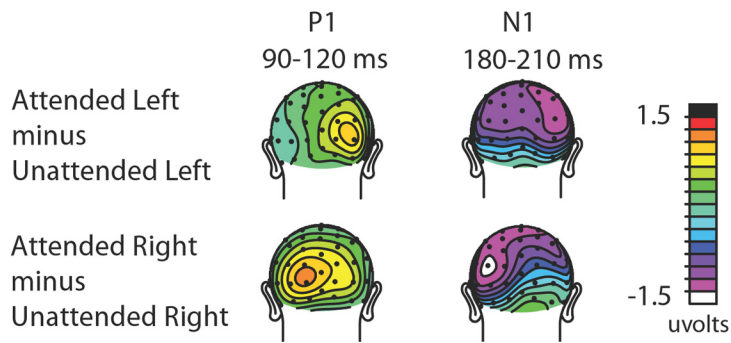


Figure 3. Visual attention ERP effects from the visual-only (VO) condition. Topographic distributions of the attention effects (attended minus unattended) for the P1 (90–120 ms) and N1 (180–210 ms) sensory components are shown separately for left-visual-field stimuli (top) and right-visual-field stimuli (bottom). As can be seen from the distributions, when the visual stimuli on the left were attended, they elicited right-sided (i.e., contralateral) P1 and N1 attention effects over occipital cortex. When the visual stimuli in the right visual field were attended, the P1 and N1 attention effects were correspondingly observed over the left visual cortex.

tone elicited in the context of occurring synchronously with an attended lateral visual stimulus from a different location. An analogous subtraction was performed to extract the ERP to the central tone when it occurred with an unattended lateral visual stimulus. Comparing these two extracted ERP responses should reveal any differences in auditory activity derived from the differential visual attention, reflecting the spreading of attention from the visual event to the synchronous auditory stimulus (Busse et al., 2005). In the simultaneous condition here, we replicated our previously reported pattern of multisensory-attentional-spread activity appearing as in Busse et al. (2005) as a long-lasting, frontocentral, negative-polarity wave from ~200 to ~700 ms ($F_{(1,17)} = 16.80, p < 0.001$, at sites Fz, FCz, FC1, and FC2; Fig. 4), elicited by the central tones occurring simultaneously with an attended lateral visual stimulus relative to an unattended one (see Table 1 for detailed statistics). The ERPs at time points before and beyond the time period of 200–700 ms did not differ significantly (all p values > 0.05). Because the spread-of-attention effect seemed to have a particularly anterior distribution, the additional, somewhat more anterior sites of Fp1m, Fp2m, and Fpz (Fig. 5) were also tested for significant differences between the attended-visual and unattended-visual multisensory conditions from 200 to 700 ms, and indeed were also highly significant ($F_{(1,17)} = 15.01, p = 0.001$).

Tone-delayed-by-100 ms condition (delay-100)

In the delay-100 condition, as indicated in Figures 4 and 5, the onset of the spreading-attention effect was shifted in time by ~100 ms, with the apparent onset of the late negative wave starting at 300 ms, rather than 200 ms. Testing the same frontal-central sites (Fz, FCz, FC1, and FC2; Fig. 5) as for the simultaneous condition revealed a significant effect of condition (attended-visual vs unattended-visual) from 300 to 800 ms ($F_{(1,17)} = 6.71, p < 0.05$; Table 1). Importantly, no effects of multisensory attentional context were found before 300 ms, supporting the presence of a 100 ms temporal shift for the attention-spreading effect when the auditory stimulus was delayed by 100 ms. As with the simultaneous condition, an additional set of more anterior electrodes (Fp1m, Fp2m, and Fpz) was also tested from 300 to 800 ms and, as above, a significant effect of multisensory attentional context was also observed over these channels ($F_{(1,17)} = 8.57, p < 0.01$).

Tone-delayed-by-300 ms condition (delay-300)

Using the same subtractive methods, the delay-300 condition was examined to look at the effects of the spreading of attention to the centrally presented auditory tone delayed by this greater interval. As indicated in Figures 4 and 5, the late negative wave observed in the simultaneous and delay-100 conditions was essentially eliminated for the delay-300 condition. The analyses revealed no 50 ms time periods that had significant attentional-spread activity between 500 and 1000 ms (Table 1), nor earlier or later; however, an analysis of the more anterior electrodes did reveal a small significant effect of the spreading of attention if taken across the entire time range (500–1000 ms: $F_{(1,17)} = 6.00, p < 0.05$; Figs. 4, 5).

Although the analyses described above show clear differences between the extracted auditory responses in the simultaneous, delay-100, and delay-300 conditions as a function of whether the accompanying lateral visual stimulus was attended versus unattended, we wanted to determine whether or not the SOA effects observed were present in both the attended-visual and unattended-visual conditions, but just larger in the attended condition, or whether they were only present in the attended condition. To assess this, we conducted two additional ANOVAs of the extracted auditory activity, separately for when the accompanying lateral visual stimulus was attended and for when it was unattended, with SOA as the main factor. The data that went into these ANOVAs were the mean amplitude values across the frontocentral ROI of sites Fz, FCz, FC1, and FC2, extracted from 200 to 700 ms for the simultaneous condition, from 300 to 800 ms for the delay-100 condition, and from 500 to 1000 ms for the delay-300 condition (i.e., after subtracting off the corresponding visual-only responses). For the attended-visual-stimulus conditions, this analysis revealed a clear main effect of the SOA, showing that there were clear differences present in the extracted auditory activity as a function of the relative delay of the onset of the auditory stimulus when the accompanying lateral visual stimulus was attended ($F_{(2,34)} = 3.28, p = 0.05$). In contrast, the corresponding ANOVA across the same channels and time periods for the extracted auditory activity for the unattended-visual-stimulus conditions showed no effect of SOA ($F < 1$), indicating that when the accompanying lateral visual stimulus was unattended, there were no significant differences in the response to the auditory stimulus as a function of its relative timing.

Distribution comparisons

A close inspection of the topographic maps (Fig. 5) suggests that there was a change in the distribution for the spreading-of-attention effect in the earliest time range of activity in the simultaneous versus the delay-100 conditions, with the simultaneous condition showing an initial period (latency 200–250 ms) of frontocentral activity, which then shifted to a more anterior position a short time later (i.e., shifting anteriorly at ~250–300 ms poststimulus onset). In contrast, the effect in the delay-100 condition appeared to lack the early frontocentral effect, having the more anterior distribution across its entire duration, suggesting the presence of an additional early

source in the simultaneous conditions that was not present in the delay-100 condition.

To determine whether this apparent distributional difference was statistically significant, we analyzed data from the 20 most anterior electrodes in different time windows. Using these electrodes, we vector-scaled the data using the McCarthy and Wood (1985) approach and then determined whether any time window-by-electrode interactions existed, which, if present, would indicate the presence of a significant shift in the distribution between those two time periods. Such a significant shift in distribution was indeed observed when comparing the initial onset of the negativity (200–250 ms) with a later portion of this long-lasting negative wave (300–350 ms; $F_{(9,153)} = 3.33$, $p = 0.001$), with the activity being more anterior in the later time window. Moreover, the delay-100 condition appeared to lack this initial, more central distribution. This was examined statistically by testing the initial period of the spreading-of-attention effect, again using the 20 most anterior electrodes, for the delay-100 condition (300–350 ms) versus the initial phase of the simultaneous condition (200–250 ms), which also revealed a significant difference in distribution ($F_{(9,153)} = 3.97$, $p = 0.0001$). In addition, the distribution of the delay-100 condition in its initial effect period from time 300 to 350 ms did not differ from the distribution of the simultaneous condition in the same (300–350 ms) latency ($F < 1$), both being the more anterior distribution. These results thus provide further converging evidence that there was an additional early source present in the simultaneous attentional-spreading activity that was not present in the delay-100 condition (Fig. 5).

Spatial shifts and ventriloquism

Ventriloquism is defined as a shift in the perceived location of a sound toward a simultaneous visual stimulus occurring in a different location (Bertelson and Radeau, 1981). Here, the tones were always centrally presented and were task irrelevant, with the attended and unattended visual stimuli being lateralized. If the perceived location of the tones was shifted in position toward the simultaneous visual stimulus when the latter was attended (as opposed to unattended), then there should be a lateral shift of the representation of the auditory stimulus in the brain toward the side contralateral to the attended visual stimulus, as observed previously in an explicit localization task (Bonath et al., 2007)—that is, the spreading-of-attention activity should be shifted contralaterally. To determine whether this occurred, the spreading-of-attention activity for each of the SOA conditions was analyzed separately for when the visual stimulus was on the left versus on the right (Fig. 6). In the top panel, the unattended simultaneous condition for left visual stimuli

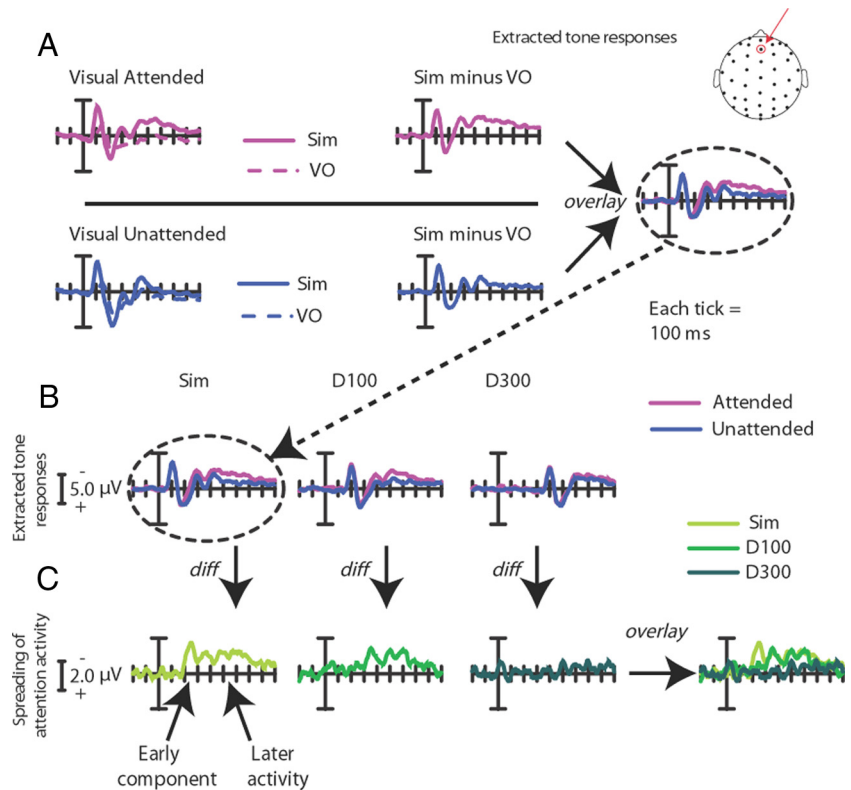


Figure 4. Multisensory spread-of-attention activity. The difference wave contrasts shown here allow for the removal of ERP activity associated with the pure sensory visual or attentional visual effects, isolating the extracted auditory ERP under each of the different multisensory attentional contexts. Note the change in scale in **A** and **B** versus **C**. All traces are for frontal site Fz (top). **A**, Extracted auditory ERP responses for the simultaneous condition. These are derived by calculating the ERP difference waves for the attended-visual simultaneous condition (Sim) minus the attended-visual visual-only condition (VO; top), and for the corresponding unattended-visual simultaneous minus unattended-visual visual-only conditions (bottom). These difference waves obtained for the extracted activity to the central tone occurring in the presence of an attended lateral visual stimulus and in the presence of an unattended lateral visual stimulus were compared (right). **B**, Extracted auditory ERP responses for all the SOA conditions. Responses for the same central auditory stimuli occurring in the context of the attended versus unattended lateral visual stimulus were calculated for each condition (simultaneous, delay-100, and delay-300). Each of these extracted-ERP difference waves was obtained in the same manner that is shown in **A** for the simultaneous condition. **C**, The difference waves between the extracted auditory responses, shown in **B** above, when they occurred in the context of an attended minus an unattended lateral visual stimulus for each of the three SOA conditions, overlaid. The simultaneous condition showed the greatest attentional difference, with the delay-100 condition showing a slightly diminished attentional difference, shifted in time by ~ 100 ms. In sharp contrast, the delay-300 displayed little differences between the extracted auditory response in the context of the attended versus the unattended visual stimuli, thus showing little of the attentional spreading effect.

was subtracted from the attended simultaneous condition for left visual stimuli (all conditions having already subtracted the respective visual-only ERP responses), analyzed for the time period of 200–250 ms. This subtraction revealed that the initial attentional-spread neural activity in the simultaneous conditions was indeed shifted toward the side contralateral to the attended visual stimulus. Correspondingly, the analogous analysis for auditory tones occurring with an attended right stimulus revealed a shift in the opposite direction (Fig. 6). As observed in the middle and bottom panels of Figure 6, this same subtraction conducted for the corresponding initial time period in the delay-100 and delay-300 conditions (from 300 to 350 and 500 to 550 ms, respectively) showed no such lateralization for the processing of the extracted responses to the auditory stimuli.

To statistically assess this effect, we performed an ANOVA of the activity in the time period from 200 to 250 ms for the simultaneous condition over the frontocentral sites C1a, C1p, C5a, C2a, C2p, and C6a, using the factors of stimulus location, hemisphere (electrode location), and attention. This analysis confirmed a significant interaction ($F_{(1,17)} = 4.58$, $p < 0.05$) between

Table 1. Statistical results for the main effect of condition for attended-visual versus unattended-visual extracted auditory activity across the frontal sites Fz, FCz, FC1, and FC2, across the entire time period in 50 ms windows for simultaneous, delay-100, and delay-300 conditions

Time window	Attended versus unattended							
	df	Simultaneous		Delay-100		Delay-300		
		F	p	F	p	F	p	
150–200 ms	1,17	<0.1	NS (0.90)	<0.1	NS (0.78)	0.12	NS (0.74)	
200–250 ms	1,17	18.83	0.0004	1.25	NS (0.28)	0.44	NS (0.52)	
250–300 ms	1,17	14.2	0.002	1.86	NS (0.20)	0.2	NS (0.66)	
300–350 ms	1,17	7.32	0.02	15.12	0.001	<0.1	NS (0.77)	
350–400 ms	1,17	8.99	0.008	10.42	0.0049	0.58	NS (0.46)	
400–450 ms	1,17	4.98	0.04	1.84	NS (0.20)	<0.1	NS (0.87)	
450–500 ms	1,17	6.31	0.02	7.7	0.01	0.62	NS (0.40)	
500–550 ms	1,17	12.44	0.003	12.96	0.002	1.57	NS (0.20)	
550–600 ms	1,17	18.4	0.0005	8.52	0.01	0.82	NS (0.40)	
600–650 ms	1,17	13.55	0.002	7.39	0.01	1.82	NS (0.20)	
650–700 ms	1,17	8.42	0.01	2.17	NS (0.20)	4.22	NS (0.06)	
700–750 ms	1,17	5.77	0.03	1.77	NS (0.20)	2.11	NS (0.20)	
750–800 ms	1,17	3.42	NS (0.08)	0.98	NS (0.30)	2.22	NS (0.20)	
800–850 ms	1,17	2.14	NS (0.16)	1.08	NS (0.30)	3.87	NS (0.07)	
850–900 ms	1,17	2.6	NS (0.13)	<0.01	NS (0.90)	3.44	NS (0.08)	
900–950 ms	1,17	1.18	NS (0.29)	0.17	NS (0.68)	0.62	NS (0.40)	
950–1000 ms	1,17	0.52	NS (0.48)	0.27	NS (0.61)	0.2	NS (0.60)	
1000–1050 ms	1,17	0.16	NS (0.69)	0.11	NS (0.74)	0.4	NS (0.50)	

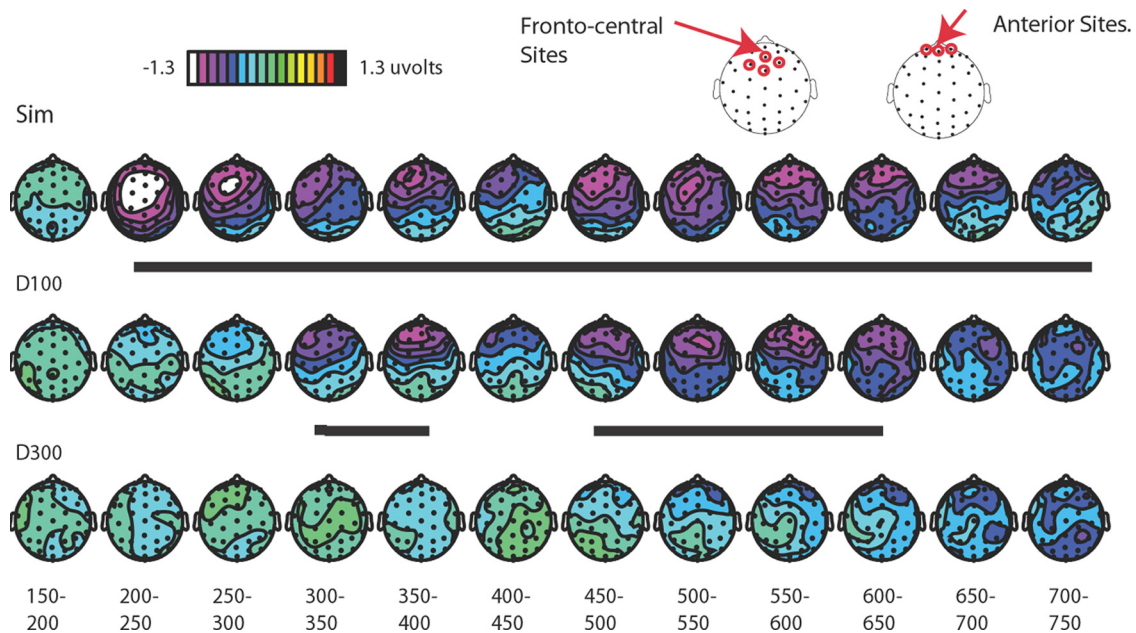


Figure 5. Topographic voltage distributions of the multisensory attentional difference waves. These distributions are displayed for the difference waves shown in Figure 4C, plotted in 50 ms bins. The multisensory attention effect is observed at frontal-central sites, maximally for the simultaneous condition, shifted in time by ~100 ms and slightly diminished for the delay condition, and essentially abolished in the delay-300 condition. Periods of significant differential activity between the attended and unattended conditions are underlined in black. Sites indicated at the top are those over which statistics were run.

these factors, attributable to the attentional-spreading effect being shifted to the side contralateral to the visual stimulus. There was also a significant lateralization interaction from 250 to 300 ms ($F_{(1,17)} = 7.36, p < 0.05$) and 300 to 350 ms ($F_{(1,17)} = 417.09, p < 0.001$); however, these later interactions were driven by shifts toward the side contralateral to the right attended visual stimulus, with no shifts toward the side contralateral to the left visual stimulus (p values > 0.1 for attended left minus unattended left on right versus left channels). No such significant interactions were observed for the delay-100 and delay-300 conditions, analyzed in the corresponding initial time windows for the effects in those conditions (300–350 ms and

500–550 ms, respectively), nor any other time windows for those conditions (F values < 1).

Discussion

This study is the first to provide a clear dissociation between the multisensory linking of the temporal and spatial aspects of the auditory and visual components of a multisensory stimulus, reflected by the spreading of attention across a multisensory object. While it is apparent that the principles of sensory integration are fundamental to the successful linking of multisensory information (Stein and Stanford, 2008), to date the degree to which the

temporal factors can interact with the spatial ones, and how these are modulated by attention, has not been much explored. Here, using an attentional manipulation and obtaining both neural and behavioral measures, we provide a new account of the spatial and temporal linking of auditory and visual stimuli, summarized in Table 2.

First, as shown in the separate behavioral sessions, participants were likely to judge our auditory and visual stimuli as occurring simultaneously when they were either actually presented simultaneously or when the auditory was delayed by 100 ms, but not when it was delayed by 300 ms, thus indicating the time window over which the stimuli are perceptually linked from a temporal standpoint. Second, in line with these behavioral findings, the neural (EEG) measures indicated that attention spread from the visual to the auditory modality when the stimuli were simultaneous or when the tone was delayed by 100 ms, but not when delayed by 300 ms. Therefore, it appears that for attention to spread successfully, the stimuli need to be temporally linked within the classic time window of audiovisual perceptual integration (Meredith et al., 1987; Schneider and Bavelier, 2003; Zampini et al., 2005; van Wassenhove et al., 2007), or, conversely, for the stimuli to be temporally linked, attention needs to spread between them. While design limitations precluded the determination of any potential modulation in the unattended channel compared with baseline, significant differences between the extracted auditory activity for the different SOA conditions were only present for attended-visual-stimulus trials, and not for unattended-visual-stimulus ones, suggesting that only attended stimuli are differentially processed and linked under our temporal manipulation. Finally, only when the auditory stimuli were presented simultaneously with the lateral visual stimulus were they spatially linked, as indicated by a shift in the neural response to the central tone when it occurred with an attended versus an unattended lateralized visual stimulus (cf. Bonath et al., 2007). Delaying the tone by even 100 ms, while still leading to the temporal linking of the stimuli and a robust spreading of attention, did not lead to any neural reflection of spatial linking, as reflected by the lack of a contralateral shift of neural processing. Importantly, because of the design of the present paradigm, the spatial linking and neural shift effects that were observed in the simultaneous conditions occurred directly as a result of a cross-sensory attentional manipulation (i.e., deriving from the same auditory stimulus occurring with an attended versus an unattended lateral visual stimulus), underscoring the importance of attention in the ventriloquism effect and in the processes underlying the spatial and temporal linking of multisensory stimulus components.

Temporal linking of multisensory stimuli

There is considerable behavioral evidence suggesting that at SOAs of increasing disparity, multisensory stimuli are more likely

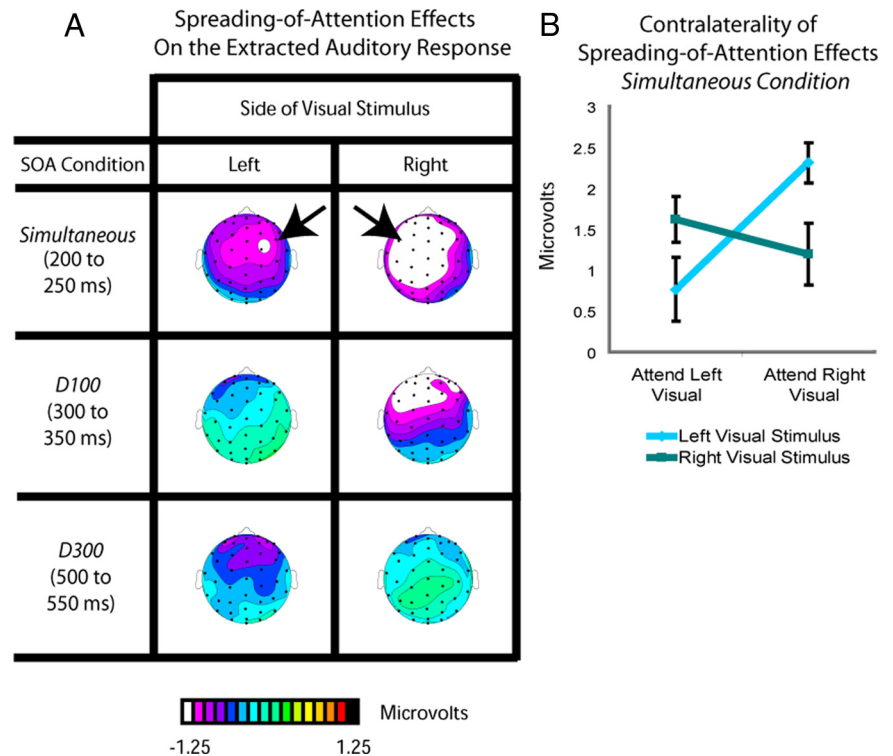


Figure 6. Distributional differences of extracted auditory response as a function of SOA and visual-stimulus laterality. **A**, Topographic voltage distributions for the spreading-of-attention effects on the extracted auditory response for each SOA condition, shown separately for the left and right visual stimuli. These spreading-of-attention effects were derived as in Figure 4 from difference waves between the auditory ERPs extracted from when the associated visual stimulus was attended versus when it was unattended (i.e., as in Fig. 5, the responses to the respective visual-only stimuli have already been subtracted). Clear differences in laterality for the spreading-of-attention activity were observed only in the simultaneous condition, and only during the early phases of that activity. **B**, Mean amplitude values over the left and right frontocentral ROIs (for right and left visual stimuli, respectively) for the spreading-of-attention activity for the centrally presented tones in the simultaneous condition, shown separately for when the associated attended visual stimulus was on the left or on the right. The plot underscores the interaction between attention and laterality observed for this initial attentional-spreading activity in the simultaneous condition, which was the only condition for which this early contralateral effect was significantly present.

Table 2. Summary of experimental results showing the dissociation between the temporal and spatial linking of multisensory stimuli and the interaction of attention with these linking processes

Stimulus	Neural	Perception
Simultaneous auditory and visual stimuli	→ Spatial Linking (200–250 ms) Temporal linking (200–700 ms) Spreading of attention (200–700 ms)	→ Spatial shift (Ventriloquism) Judged as simultaneous
Auditory tone delayed by 100 ms compared with visual	→ Temporal linking (300–800 ms) Spreading of attention (300–800 ms)	→ No spatial shift Judged as simultaneous
Auditory tone delayed by 300 ms compared with visual	→ No spatial linking No temporal linking No attentional spreading	→ No spatial shift Not judged as simultaneous

to be perceived as separate (Spence et al., 2001; Schneider and Bavelier, 2003; Zampini et al., 2005), as observed here in our simultaneity-judgment task. Further, in multisensory speech identification, a temporal separation of more than ~150 ms generally has little or no audio-visual benefit behaviorally when compared with visual alone (McGrath and Summerfield, 1985),

whereas there is a clear benefit at SOAs of less than this separation. Neurally, semantically unrelated stimuli (e.g., a tone and a light) are linked together, as indexed by enhanced firing to the multisensory stimulus, but only when they occur within the temporal window of integration (Meredith et al., 1987).

Consistent with this previous work, our findings indicate that simple multisensory stimulus components will be perceptually linked when presented within the temporal window of integration. More importantly here, however, we demonstrate that attention will spread from one modality to another only when stimuli are presented within this temporal window of integration, thereby illustrating the correspondence between the temporal window of integration and the temporal window over which attention will spread across modalities. Indeed, the present data suggest the intriguing hypothesis for the key role played by attention, and by its striking ability to spread across sensory modalities, to potentially aid in (or be necessary for) the temporal linking of the component features of multisensory stimulus input into a perceptual whole. One might speculate further that it may be this sort of attentional spread that underlies, or at least contributes to, the perceptual linking of the various features of any multi-featured object (Schoenfeld et al., 2003).

Spatial linking of multisensory stimuli

Another important new finding here is that of the tighter temporal constraints that appear to be required for the spatial linking of the different multisensory components. Only in the simultaneous condition, at the onset of the negative-polarity wave reflecting the attentional-spreading activity for the centrally presented tone, did we observe an additional ERP component showing a lateral shift to the side contralateral to the visual stimulus. The location and timing of this lateralized neural activity (centrally/frontocentrally distributed, occurring at ~200–250 ms) is very similar to that found in an explicit auditory localization task by Bonath et al. (2007). In that study, on trials in which the percept of the spatial location of the auditory stimulus was shifted toward the visual, there was a corresponding lateralized shift in the distribution of the ERP activity contralateral to the location of the visual stimulus, with this activity being modeled as arising from auditory cortex (Bonath et al., 2007). The fMRI part of our previous study confirmed the presence of spreading-of-attention activity in auditory cortex (Busse et al., 2005), where similar regions in the planum temporale are involved in discriminating the spatial location of sound (Deouell et al., 2007).

In the present study, another important aspect of the ventriloquist-related finding was that the observed neural processing shift occurred directly as a result of an attentional manipulation, emerging as a difference for identical tones that occur simultaneously with an attended versus an unattended visual stimulus. Moreover, the effect occurred only in the simultaneous condition, was present only in the initial 50–100 ms phase of the activity, and was elicited for auditory stimuli that were completely task irrelevant. While the delay of 100 ms allowed attention to still spread from the visual modality to the auditory modality, as reflected by the elicitation of the sustained negative-polarity ERP wave, and also resulted in the stimuli being still judged as synchronous in the separate behavioral experiment, this temporal offset appeared to be enough to abolish the lateral neural-processing shift associated with a ventriloquist effect. This neural result is consistent with previous behavioral studies reporting that increasing the audio-visual temporal separation reduces the perceived location shift of the auditory stimulus toward the visual, compared with stimuli presented simultaneously or

delayed by only 50 ms (Slutsky and Recanzone, 2001; Lewald and Guski, 2003). The present study provides the first electrophysiological evidence for the temporal limits of the neural processes that lead to the ventriloquist illusion.

Further, the present study emphasizes the important role of visual attention for this lateralization effect to occur, since it occurred directly as a result of the attentional manipulation. Although previous behavioral studies have suggested that the ventriloquist illusion is preattentive and not influenced by attention (Bertelson et al., 2000), the present findings argue strongly against such a conclusion, with explicit neural evidence showing a direct modulation of the spatial linking of auditory and visual stimuli as a function of attention. The present findings showing the key role of attention on multisensory integration are in line with other recent evidence suggesting that other multisensory illusions, such as the sound-induced extra-flash visual illusion (Shams et al., 2000), can be modulated by attention (Mishra et al., 2010).

To summarize, we show that visual attention can spread robustly across both modality and space to a task-irrelevant and spatially separated auditory tone when it occurs within the temporal window of integration, with this spread essentially being eliminated for SOAs outside that window. Further, when the auditory stimulus is delayed in time relative to the visual event, but is still within the temporal window of integration, the visual attention will spread (albeit slightly attenuated), and this spreading will be delayed by the delay in the onset of the stimulus. Finally, only when auditory and visual stimuli occur in close temporal proximity is there a spatial linking of the unisensory components. In particular, only under these circumstances was there a shift of the auditory neural processing to the side contralateral to the visual stimulus, consistent with a ventriloquist-like perceptual shift of the centrally presented auditory stimulus toward the visual. Moreover, all of these cross-modal effects on the processing of task-irrelevant auditory stimulus occurred directly as a function of whether the synchronous or near-synchronous visual stimulus was or was not attended, underscoring the fundamental role of attention in these multisensory integration processes (Talsma et al., 2010). These results thus shed new light on the temporal and spatial constraints by which the various unisensory components of multisensory stimuli are linked together into a perceptual whole, and the way in which attention modulates these stimulus-linking processes.

References

- Bertelson P, Radeau M (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys* 29:578–584.
- Bertelson P, Vroomen J, de Gelder B, Driver J (2000) The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept Psychophys* 62:321–332.
- Bolognini N, Leo F, Passamonti C, Stein BE, Ladavas E (2007) Multisensory-mediated auditory localization. *Perception* 36:1477–1485.
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA (2007) Neural basis of the ventriloquist illusion. *Curr Biol* 17:1697–1703.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. *Proc Natl Acad Sci U S A* 102:18751–18756.
- Deouell LY, Heller AS, Malach R, D'Esposito M, Knight RT (2007) Cerebral responses to change in spatial location of unattended sounds. *Neuron* 55:985–996.
- Egley R, Driver J, Rafal RD (1994) Shifting visual-attention between objects and locations—evidence from normal and parietal lesion subjects. *J Exp Psychol Gen* 123:161–177.
- Fiebelkorn IC, Foxe JJ, Molholm S (2010) Dual mechanisms for the cross-

- sensory spread of attention: how much do learned associations matter? *Cereb Cortex* 20:109–120.
- Gondan M, Niederhaus B, Rösler F, Röder B (2005) Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Percept Psychophys* 67:713–726.
- Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003) Visual localization ability influences cross-modal bias. *J Cogn Neurosci* 15:20–29.
- Hillyard SA, Anllo-Vento L (1998) Event-related brain potentials in the study of visual selective attention. *Proc Natl Acad Sci U S A* 95:781–787.
- Keetels M, Vroomen J (2005) The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Exp Brain Res* 167:635–640.
- Lewald J, Guski R (2003) Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Res Cogn Brain Res* 16:468–478.
- Martinez A, Ramanathan DS, Foxe JJ, Javitt DC, Hillyard SA (2007) The role of spatial attention in the selection of real and illusory objects. *J Neurosci* 27:7963–7973.
- 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.
- McGrath M, Summerfield Q (1985) Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *J Acoust Soc Am* 77:678–685.
- Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol* 56:640–662.
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215–3229.
- Mishra J, Martinez A, Hillyard SA (2010) Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *J Cogn Neurosci* 22:1714–1729.
- Molholm S, Martinez A, Shpaner M, Foxe JJ (2007) Object-based attention is multisensory: coactivation of an object's representations in ignored sensory modalities. *Eur J Neurosci* 26:499–509.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Schneider KA, Bavelier D (2003) Components of visual prior entry. *Cogn Psychol* 47:333–366.
- Schoenfeld MA, Tempelmann C, Martinez A, Hopf JM, Sattler C, Heinze HJ, Hillyard SA (2003) Dynamics of feature binding during object-selective attention. *Proc Natl Acad Sci U S A* 100:11806–11811.
- Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. *Nature* 408:788.
- Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12:7–10.
- Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. *J Exp Psychol Gen* 130:799–832.
- Spence C, Baddeley R, Zampini M, James R, Shore DI (2003) Multisensory temporal order judgments: When two locations are better than one. *Percept Psychophys* 65:318–328.
- Stein BE, Meredith MA (1993) *The merging of the senses*. Cambridge, MA: MIT.
- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9:255–266.
- Stone JV, Hunkin NM, Porrill J, Wood R, Keeler V, Beanland M, Port M, Porter NR (2001) When is now? Perception of simultaneity. *Proc Biol Sci* 268:31–38.
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 14:400–410.
- 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.
- van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45: 598–607.
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76:1246–1266.
- Zampini M, Guest S, Shore DI, Spence C (2005) Audio-visual simultaneity judgments. *Percept Psychophys* 67:531–544.