

# The spread of attention across modalities and space in a multisensory object

Laura Busse, Kenneth C. Roberts, Roy E. Crist, Daniel H. Weissman, and Marty G. Woldorff\*

Center for Cognitive Neuroscience and Department of Psychiatry, Duke University, Durham, NC 27708

Communicated by Dale Purves, Duke University Medical Center, Durham, NC, October 7, 2005 (received for review December 20, 2004)

**Attending to a stimulus is known to enhance the neural responses to that stimulus. Recent experiments on visual attention have shown that this modulation can have object-based characteristics, such that, when certain parts of a visual object are attended, other parts automatically also receive enhanced processing. Here, we investigated whether visual attention can modulate neural responses to other components of a multisensory object defined by synchronous, but spatially disparate, auditory and visual stimuli. The audiovisual integration of such multisensory stimuli typically leads to mislocalization of the sound toward the visual stimulus (ventriloquism illusion). Using event-related potentials and functional MRI, we found that the brain's response to task-irrelevant sounds occurring synchronously with a visual stimulus from a different location was larger when that accompanying visual stimulus was attended versus unattended. The event-related potential effect consisted of sustained, frontally distributed, brain activity that emerged relatively late in processing, an effect resembling attention-related enhancements seen at earlier latencies during intramodal auditory attention. Moreover, the functional MRI data confirmed that the effect included specific enhancement of activity in auditory cortex. These findings indicate that attention to one sensory modality can spread to encompass simultaneous signals from another modality, even when they are task-irrelevant and from a different location. This cross-modal attentional spread appears to reflect an object-based, late selection process wherein spatially discrepant auditory stimulation is grouped with synchronous attended visual input into a multisensory object, resulting in the auditory information being pulled into the attentional spotlight and bestowed with enhanced processing.**

auditory | event-related potential | functional MRI | ventriloquism | visual

Attention allows us to dynamically select and enhance the processing of stimuli and events that are most relevant at each moment. Directing attention to a stimulus leads to lower perceptual thresholds, faster reaction times (RTs), and increased discrimination accuracy (1–2). The physiological basis for these perceptual benefits involves enhanced neural activity in response to the attended stimulus, as has been shown by a variety of brain activity measures (e.g., refs. 3–5).

Attention can be directed to a spatial location (2–3, 6–7), enhancing the processing of all stimuli occurring at that location, or to a particular stimulus feature (e.g., the color red, a certain tonal frequency), resulting in preferential processing of that feature independent of its spatial location (8–9). Furthermore, it has been proposed that attention can also act on whole objects, such that when attention is directed to one part of an object, the other components of the same object automatically receive enhanced processing (10–12).

So far, studies of object-based attentional selection have focused on the visual modality. In these experiments, subjects are typically cued to direct attention to a spatially defined part of one of two presented objects. The results show that performance is enhanced for stimuli at an unexpected location within the cued object compared to stimuli at an equidistant location within the uncued object (same-object advantage). This finding has been interpreted as reflecting an automatic spread of

attention through the cued object (reviewed in ref. 13). Recently, a functional magnetic resonance imaging (fMRI) study using a variant of this paradigm reported enhancement of lower-tier visual cortical activity at retinotopic representations of both the cued and the uncued locations within the attended object, whereas retinotopic representations of locations in the uncued object were not modulated by attention (14).

In that the real world is multisensory, our performance often critically depends on our ability to attend to and integrate the various features from multisensory objects (15). Here, we investigated whether object-based attentional selection can occur with audiovisual multisensory objects, asking whether visual attention might spread across space and modality to enhance a simultaneous task-irrelevant auditory signal from an unattended location. Evidence that auditory and visual stimuli occurring in temporal synchrony, but at disparate spatial locations, are perceptually grouped into a single multisensory object has come from studies of the ventriloquism illusion (16–18). In such a situation, observers mislocalize the auditory stimulus toward the location of the visual stimulus (19–21). There are a couple of reasons why this multisensory stimulus configuration is particularly well suited for investigating whether and how attention might spread across space and across sensory modalities. First, auditory and visual information tend to be integrated into a single multisensory object as evidenced by the ventriloquism illusion. Second, because of the spatial separation of the auditory and visual stimuli, spatial attention to the visual stimulus could not explain any modulation of the auditory stimulus response.

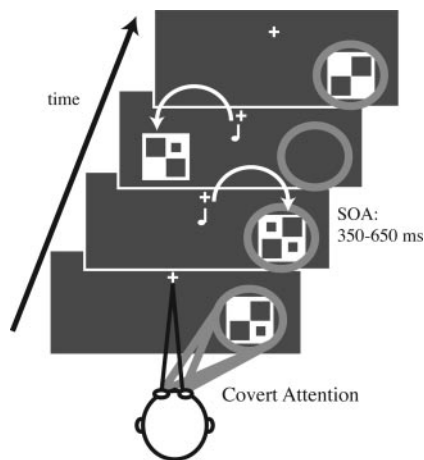
To investigate the question of multisensory, object-based spreading of attention, we combined recordings of event-related potentials (ERPs), providing high temporal resolution of any such effects, and event-related fMRI, offering high spatial resolution for localizing their neuroanatomical sources. One possible outcome of such a study would be that attention to the visual stimulus might have no effect on the processing of a simultaneous auditory stimulus from an unattended location. Indeed, some previous behavioral studies have suggested that the ventriloquism effect is not influenced by attention (22–23). However, if there were an effect of visual attention on the processing of a simultaneous auditory signal, at least two alternative mechanisms could be envisioned. On the one hand, there could be an early influence of attention after rapid integration processes, probably arising from multisensory interactions in the brainstem or lower-tier sensory cortices. This early attentional influence would be evidenced by ERP modulations as early as 20–100 ms, latencies typically found for unisensory auditory attention effects (P20–50/M20–50, N1 effect) (24–26). On the other hand, later attentional modulations would favor a model in which the system initially processes the spatially disparate visual and auditory input separately, detects their temporal coincidence, links the information together into a single multisensory object, and

Conflict of interest statement: No conflicts declared.

Abbreviations: ERP, event-related potentials; fMRI, functional magnetic resonance imaging; ROI, region of interest; RT, reaction time.

\*To whom correspondence should be addressed. E-mail: woldorff@duke.edu.

© 2005 by The National Academy of Sciences of the USA



**Fig. 1.** Stimuli and task. In different runs, subjects covertly directed attention to the stream of visual stimuli on one side of the monitor while ignoring all visual stimuli on the opposite side and all auditory stimuli. Visual stimuli were unilateral streams of flashing checkerboards randomly presented on each side at stimulus onset asynchronies (SOAs) of 350–650 ms. The subjects' task was to detect target stimuli (checkerboards with two dots) occurring infrequently (14%) at the attended location. Half of the visual stimuli were accompanied by a simultaneous task-irrelevant tone presented centrally.

ultimately modulates the processing of all its components, including the spatially discrepant tones. In either case, the corresponding fMRI data should reveal that multisensory, object-based spreading of attention from visual to auditory stimuli would result in enhanced activity in auditory cortex. Furthermore, such auditory cortex activity should also be reflected in a fronto-central ERP scalp topography because of a major contribution from a dipolar pair of sources in the superior temporal plane pointing upward and slightly forward (27).

The results show that attention can indeed spread across modalities and across space as evidenced by enhanced neural responses to the task-irrelevant auditory component of such a multisensory audiovisual object, including in modality-specific auditory sensory cortex. The latency of the attentional modulation clearly favors the notion of a late, object-based, attentional selection process.

## Methods

**Participants. ERPs.** Seventeen adults participated in the ERP experiment. Six subjects were excluded due to poor performance (four subjects) or the loss of >50% of trials due to physiological artifacts (two subjects), leaving 11 subjects in the final analysis (five male; mean age, 24 years).

**fMRI.** Twenty-six adults participated in the fMRI experiment. Seven subjects were excluded due to technical problems with the stimulus equipment, leaving 19 in the final analysis (nine male, mean age, 25 years).

The study protocol was approved by the Duke University Health System Institutional Review Board, and written informed consent was obtained from all participants.

**Stimuli and Task.** We investigated the brain activation patterns of object-based attentional selection by using a specially designed attentional stream paradigm. In two different conditions, subjects fixated on a central point while covertly directing visual attention to either the left or right side of the monitor. Streams of brief (40 ms), unilateral, visual stimuli (checkerboards containing 0, 1, or 2 dots) were rapidly presented [stimulus onset asynchronies (SOAs) = 350–650 ms] to the lower left and right visual quadrants (Fig. 1). Subjects pressed a button with their right index finger upon detecting infrequent (14% probability) target stimuli in the stream at the attended location. Targets were checkerboards containing 2

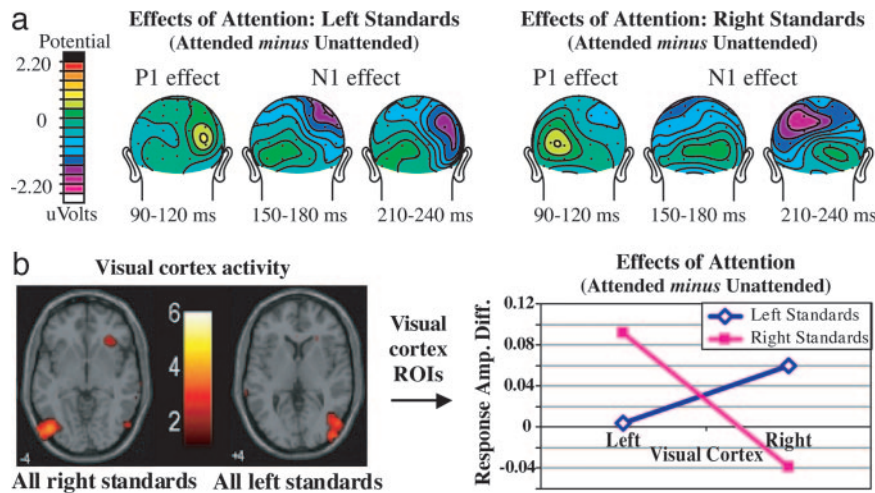
dots; those with only one or no dots were “standards.” Subjects were instructed to ignore all stimuli (both standards and targets) on the unattended side. RTs and accuracy of responses to targets were recorded for all runs. Target difficulty was titrated for each subject so that detection required highly focused attention but could be performed at ≈80% correct. This adjustment was accomplished by slightly changing the contrast and/or the size of the dots.

Half of these lateral visual stimuli were accompanied by a task-irrelevant simultaneous tone pip (pitch 1200 Hz; intensity 60 dB SL; duration 25 ms, including 5-ms rise and fall periods). In the ERP experiment, the tones were presented centrally from behind the monitor. In the fMRI experiment, the tones were delivered binaurally via headphones. To better match the ERP auditory stimulation, the fMRI tones were processed by a 3D sound program (Q Creator, Q Sound Labs, Inc.) to generate the impression that they originated from ≈1 m in front of the subject. In both settings, the auditory stimuli played alone were perceived as centrally presented from directly in front of the subject. When these tones occurred simultaneously with the lateralized visual stimuli, however, the ventriloquism effect shifted the perceptual origin of the tones toward the locations of those visual stimuli. For the experiment, however, subjects were naïve concerning the purpose of the tones and were instructed to ignore them.

An additional 1/5 of the trials were “no-stim” events (points in time in the trial sequence randomized like real stimulus events but in which no stimulus actually occurs), allowing for adjacent-response overlap removal in fast-rate event-related fMRI (28–29) and ERP (30–31) designs. The various trial types (1/5 no-stims, 1/5 left visual with tone, 1/5 left visual alone, 1/5 right visual with tone, and 1/5 right visual alone) within each run were first-order counterbalanced, i.e., each stimulus type was preceded and followed by all stimulus types equally often. Single runs lasted ≈2 min. The order of the attention conditions (attend visual left and attend visual right) was randomized across the runs.

**ERP Recordings and Analysis.** The electroencephalogram (EEG) was recorded continuously from 64 electrodes in a customized elastic cap (Electro-Cap International) with a bandpass filter of 0.01–100 Hz and a sampling rate of 500 Hz (SynAmps, Neuroscan). Fixation and eye movements were monitored with both electro-oculogram (EOG) recordings and a zoom-lens video camera. Artifact rejection was performed off-line by discarding EEG/EOG epochs contaminated by eye movements, eye blinks, excessive muscle activity, drifts, or amplifier blocking. ERP averages to the various trial types were extracted by time-locked averaging and then digitally low-pass filtered (<57 Hz) and rereferenced to the algebraic average of the two mastoid electrodes. Difference waves based on the direction of attention and on multisensory attentional context were calculated, as described in *Results*. The ERPs and ERP difference waves were grand-averaged across subjects. Repeated-measures analyses of variance (ANOVAs) were performed on amplitude measures of the ERP waveforms and difference waves across subjects, relative to a 200-ms prestimulus baseline.

**fMRI Recordings and Analysis.** The fMRI data were collected on a General Electric 4T scanner by using an inverse spiral pulse sequence with a repetition time (TR) of 1.5 s. Whole-head fMRI activity was recorded by using 32 slices parallel to the anterior commissure–posterior commissure line with isotropic voxels 3.75 mm on a side. The data were preprocessed (slice-time corrected, realigned, normalized to MNI space, and spatially smoothed) by using SPM99. Event-related fMRI responses to the different trial types were modeled by convolving a canonical hemodynamic response function with the onset times for each trial type. The general linear model in SPM99 was then used to estimate the magnitude of the response produced by each trial type (32). To statistically assess the response differences to stimuli between conditions (see *Results*), a region of interest (ROI) was made



**Fig. 2.** Visual attention effects on unisensory visual standards. (a) ERPs effects: Topographic plots of the attention effects in the left and right stream. For each side, the P1 effect (90–120 ms) is most prominent at contralateral occipital electrodes. It is followed by the N1 effect, first at contralateral fronto-central sites (150–200 ms) and later at parietal sites (225–250 ms). (b) fMRI effects: Event-related activation maps (Left) show the responses in contralateral visual cortex for left and right visual standards, collapsed across attended and unattended conditions. These activations were then used as ROIs for analyzing the effects of attention (i.e., response amplitude for attended vs. unattended visual standards), which revealed enhanced activity contralaterally (Right).

around a local maximum of the map defined by the same stimuli collapsed across the two conditions. Random-effects, paired *t* tests were then performed within the ROIs by using the regression coefficients for each subject.

**Behavioral Data Analysis.** Target hit rates (percent correct), false alarm rates, and RTs for correct detections of targets were computed separately for the different conditions. Only responses occurring from 200–1,000 ms after target presentation were counted as correct target detections. Paired *t* tests were performed on the RTs, hit rates, and false alarm rates between experimental conditions.

## Results

**Behavioral Data.** In both the ERP and fMRI sessions, the hit rates provided behavioral evidence for multisensory grouping of the simultaneous, but spatially discrepant, visual and auditory stimuli. Subjects detected significantly more visual targets when they were accompanied by an auditory stimulus than when they were presented alone [ERP session: 82.6% vs. 78.2%;  $t_{10} = 2.31$ ,  $P < 0.02$ ; fMRI session, 80.6% vs. 78.3%,  $t_{18} = 1.7$ ,  $P = 0.053$  (one-tailed tests)], despite the auditory stimuli being task-irrelevant. No significant RT differences were observed. No significant differences in any of the behavioral measures were found for attending left vs. right.

**Effects of Visual Attention on the Unisensory Visual Stimuli.** To confirm the effectiveness of our visual attention manipulation, we compared the visual activity to attended vs. unattended visual standards presented alone. For the ERP data (Fig. 2a), we found the characteristic attentional modulations of the visual components at contralateral electrode sites (4, 33), consisting of an initial enhancement of the scalp-positive P1 wave (P1 effect) followed by an increased negativity around the N1 latency (N1 effect) [ANOVA: three-way interaction between the factors attention, side of stimulus, and hemisphere:  $F_{1,10} = 11.9$ ,  $P < 0.01$  (P1 effect, 110–140 ms, across electrodes TO1/TO2, O1i/O2i, and P3i/P4i);  $F_{1,10} = 5.12$ ,  $P < 0.05$  (N1 effect, frontal sites, 150–170 ms, across electrodes C1a/C2a and C3a/C4a);  $F_{1,10} = 14.16$ ,  $P < 0.005$  (N1 effect, parietal sites, 225–250 ms, across electrodes P3i/P4i, P3a/P4a, and O1'/O2')]. Additionally, target stimuli in the attended stream elicited large P300 waves (latency  $\approx 300$ –600 ms) associated

with target detection, which were absent for targets on the unattended side (data not shown).

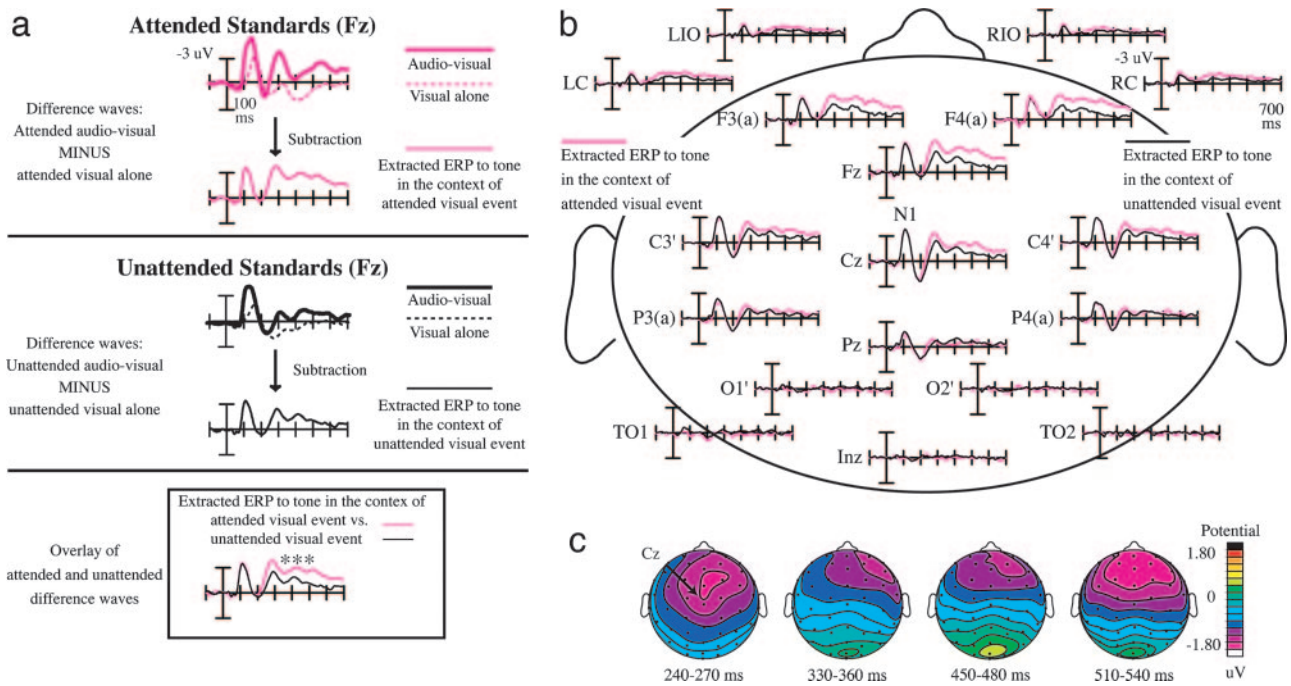
For the fMRI data, an analogous analysis was performed comparing responses to attended vs. unattended visual standards presented alone (Fig. 2b). As expected, responses to lateral visual stimuli in contralateral visual cortex were larger when those stimuli were attended vs. unattended (right standards, left visual cortex ROI:  $P < 0.01$ ; left standards, right visual cortex ROI:  $P < 0.05$ ).

Taken together, these effects of sustained visual attention demonstrate that subjects were indeed focusing their attention on the designated stream of visual stimuli while ignoring the visual events at the unattended location.

**Attention Effects on the Spatially Discrepant Auditory Stimuli.** The key comparison for assessing the spread of attention across the audiovisual multisensory objects was a contrast of responses (ERP or fMRI) elicited by tones occurring simultaneously with an attended vs. an unattended visual stimulus. Because the tones were always task-irrelevant, physically identical, and presented from a central location outside the focus of visual attention, any differences between these responses can be explained only by differential processing due to the multisensory attentional context of the tones, that is, whether they had occurred simultaneously with an attended or an unattended visual stimulus.

For the ERPs, to delineate the influence of the multisensory attentional context on the synchronous, but spatially discrepant, tones, we used a sequence of ERP subtractions. Fig. 3a shows an overview of the subtractions for frontal scalp site Fz. For ease of illustration, the responses are collapsed across sides of attention in the figure; the analyses, however, were performed on the uncollapsed data. In a first step, we subtracted the ERPs to the visual stimuli presented alone from the combined audiovisual stimuli, separately for each attention condition and side. This subtraction removes the simple visual sensory components and visual attention effects that are common to the visual standards presented alone and the visual part of the multisensory stimuli. Thus, the resulting ERP difference waves (site Fz, Fig. 3a Bottom; sites across the head, Fig. 3b) mainly reflect the brain activity elicited by the tones, including any interactions due to their occurring in the context of an attended visual standard (pink traces) vs. an unattended visual standard (black traces). As a simple proof of principle, the distribution and time course of the resulting extracted “auditory” waves (P1 at





**Fig. 3.** Multisensory object-based attention effect on the task-irrelevant midline tones, collapsed across the attend-left and attend-right conditions. (a) Sequence of ERP subtractions leading to the isolation of the multisensory attention effects for the tones, shown at frontal site Fz. (Top) Attended condition: A mixture of auditory and visual components can be seen in the ERP to the combined audiovisual events (thick trace). The subtraction of the visual-alone ERP (dotted trace) from the audiovisual ERP yields the “extracted” ERPs to the tones in the context of an attended visual stimulus (thin solid trace). (Center) Unattended condition: The analogous subtraction is performed on the unattended unisensory visual standards (dotted trace) and the unattended visual standards paired with a central tone (thick trace). The thin solid trace shows the corresponding unattended-condition difference wave of the multisensory minus unisensory-visual ERPs. (Bottom) The extracted difference waves overlaid, revealing the multisensory attention effect on the synchronous tones. An attention-related difference, a frontally distributed processing negativity, emerges  $\approx 220$  ms and lasts for hundreds of milliseconds. (b) Multisensory object-based attention effect, for a number of electrode sites, laid out as on the subject’s head. Although there is no difference between the attention conditions on the early auditory components (P20–50, N1), a sustained attentional difference starts to emerge at  $\approx 220$  ms over fronto-central and frontal sites and lasts for  $>400$  ms. (c) Topographic voltage maps for the multisensory attention effect on the spatially discrepant tones. Shown are the difference maps for tones in the context of an attended vs. unattended lateral visual stimulus (see corresponding ERP waves in b). This attentional difference is maximal over fronto-central and frontal sites, similar to that of the attention-related N1 effect and processing negativity elicited by attended auditory stimuli in auditory attention experiments.

$\approx 50$ ms, N1 at  $\approx 100$  ms, and P2 at  $\approx 180$  ms) were highly similar to unisensory auditory ERPs (34). Furthermore, in these extracted auditory ERP responses, the occipital electrodes that reflect evoked visual activity and are the primary sites for early visual attention effects showed no significant evoked activity (Fig. 3b), as expected for responses to auditory stimuli.

The statistical analyses of the extracted ERPs to tones within the context of the two visual attention conditions (Fig. 3) revealed no early attention effects for time windows between 20–50 ms (P20–50/M20–50 wave,  $P = 0.26$ ) nor at  $\approx 100$  ms (N1,  $P = 0.64$ ). A very prominent attentional modulation, however, started at  $\approx 220$  ms, consisting of a sustained, frontally distributed, processing negativity that lasted several hundred milliseconds (main effect of attention:  $F_{1,10} = 13.71$ ,  $P < 0.005$ , 220–700 ms, across electrodes Fz, Fcz, FC1, and FC2). A more detailed ANOVA across consecutive 20-ms time windows starting at 200 ms revealed a significant negativity for attended-context tones that started at 220 ms and persisted until  $\approx 700$  ms (main effect of attention:  $F_{1,10}$  ranging between 5.41 and 35.32,  $0.04 < P < 0.0001$  for the 20-ms windows across the above electrodes).

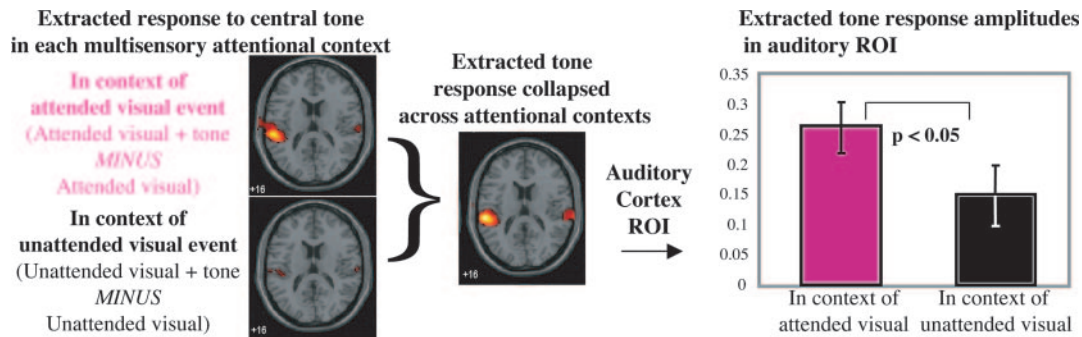
Topographic maps of voltage distributions of the difference waves between these ERP traces (i.e., the difference waves of “Extracted ERPs to tones in the context of an attended visual standard” minus “Extracted ERPs to tones in the context of an unattended visual standard”) show that the attention-related difference was maximal over fronto-central and frontal scalp regions (Fig. 3c). This scalp topography has some similarity to the voltage distribution of the auditory sensory N1 component at  $\approx 100$  ms and

to the enhanced processing negativity wave elicited by attended stimuli (at earlier latencies than seen here) in various unimodal auditory attention experiments.

For the fMRI, analogous to the ERPs, the multisensory attentional context effect for the tones was extracted by first contrasting the fMRI responses to the audiovisual events with those to the corresponding visual-alone stimuli, yielding extracted tone-related fMRI brain responses for the two attention conditions (Fig. 4). As with the ERPs, despite the tones all being physically identical, task-irrelevant, and from an unattended location, tones paired with an attended visual stimulus elicited an enhanced response, with the fMRI data showing that this specifically included enhanced activity in auditory cortex. ROI analyses of the extracted responses in auditory cortex showed that the responses were significantly greater when the auditory stimuli occurred synchronously with an attended vs. an unattended visual stimulus ( $t_{18} = 2.18$ ,  $P < 0.05$ , one-tailed; Fig. 4).

### Discussion

In this study, we investigated how visual spatial attention to visual input during multisensory stimulation affects the processing of simultaneous auditory stimuli arising from a different location. Multisensory interactions between the simultaneous auditory and visual stimuli were reflected by improved detection of visual target stimuli accompanied by a task-irrelevant auditory tone, consistent with previous studies reporting such behavioral effects (35–36). Furthermore, both the ERP and fMRI data show that attention to the visual stimulus selectively enhanced the brain’s response to a



**Fig. 4.** fMRI response showing the multisensory attentional context effects on the task-irrelevant midline tones in auditory cortex, collapsed across the attend-left and attend-right conditions. (Left) Extracted event-related tone responses, shown separately for when synchronous with an attended vs. an unattended lateral visual stimulus. (Middle) Extracted event-related tone response, collapsed across attention conditions, to obtain auditory cortex ROIs. (Right) Extracted event-related response amplitudes in these auditory cortex ROIs for each of the multisensory attention context conditions.

task-irrelevant, spatially discrepant tone when it occurred synchronously with the attended visual stimulus, despite the tone occurring in an unattended location and being task-irrelevant. In other words, neural activity elicited by identical auditory stimulation was modulated as a function of whether attention was directed toward versus away from a temporally cooccurring visual stimulus at a different location. The ERP reflection of this enhanced activity consisted of a sustained, frontally distributed brain wave that emerged rather late in processing ( $\approx 220$  ms after stimulus). The distribution of this enhanced late ERP activity for the task-irrelevant auditory stimuli was consistent with it including major contributions from auditory cortex. Moreover, the corresponding fMRI data confirmed that the additional activity elicited by the irrelevant auditory stimulus included modality-specific enhancement in auditory cortex.

These findings provide compelling evidence that the simultaneity of the visual and auditory stimuli can cause attention to spread across modalities and space to encompass the concomitant auditory signals. This cross-modal, attentional interaction effect appears to reflect a late attentional selection process by which the sensory signals from an initially unattended modality and location are grouped together with synchronous attended sensory input into a multisensory object. This grouping leads the initially unattended modality signals to be pulled into the attentional spotlight and enhanced, with the latency of the effect indicating that this attentional spread occurs relatively late in stimulus processing. Hence, analogous to object-based attentional selection in the visual modality, attention can spread across the various sensory components and even spatial locations of a multisensory object.

It could be argued that the enhanced activity for the irrelevant midline tone might not be specific for the auditory stimulus and thus does not necessarily reflect its grouping with the synchronous attended visual stimulus. In particular, the occurrence of an attended visual stimulus might lead to a brief period of general increased neural responsiveness, such that the processing of any stimulus occurring simultaneously with the attended one would tend to be enhanced. There are two lines of argument against this possibility. First, the fMRI results confirmed that the enhanced processing for the irrelevant midline tone included modality-specific enhancement of activity in auditory cortex, rather than consisting only of activity enhancement that is nonspecific with respect to modality. Second, we performed a control ERP study examining analogous visual–visual interactions with a very similar paradigm (see supporting information, which is published on the PNAS web site). As in the multisensory study, subjects attended to either a left or right stream of unilateral visual stimuli, but with half of them being accompanied by an irrelevant visual stimulus in the midline, rather than by an irrelevant auditory stimulus. Using analogous subtractions, the response to the task-irrelevant midline visual stimulus was extracted as a function of whether it occurred

simultaneously with an attended vs. an unattended lateral visual stimulus. In contrast to the multisensory case, there were no significant differences for these extracted responses (see supporting information), including at the longer latencies past 200 ms, where the robust multisensory attentional effects were seen. Taken together, these results argue that the enhanced extracted auditory response was not a result of increased nonspecific neural responsiveness, but rather a reflection of the specific grouping of the auditory stimulus with the simultaneously attended visual stimulus and the spread of attention to that auditory stimulus.

The high temporal resolution of the ERPs revealed that the object-based attentional modulation started relatively late in processing,  $\approx 220$  ms after stimulus onset. This result can be contrasted to previous results from multimodal attentional stream ERP experiments that have revealed relatively early effects of attention on stimulus processing (e.g., refs. 37–39). In these experiments, subjects were presented with streams of interspersed unisensory visual and auditory stimuli at two different locations. When subjects attended to the visual stimuli at one of the locations, the processing of the task-irrelevant auditory stimuli occurring at that same location was also enhanced. These attention effects were similar in time course and distribution to (although typically smaller than) the effects on auditory processing in conditions when subjects were attending to one auditory stream vs. another (i.e., enhancement of the auditory N1 component at  $\approx 100$  ms, followed by a fronto/ fronto-central sustained processing negativity; ref. 40). These modulations have been interpreted as reflecting the engagement of supramodal spatial attention mechanisms that encompass all stimuli occurring at the attended location.

Our present result of a late attentional selection thus differs from these prior multimodal attentional stream results. Unlike in prior studies, however, the auditory stimuli here occurred synchronously with a visual stimulus. In addition, they were never presented at the location of the focus of visual attention. Accordingly, it was impossible to foresee whether an upcoming auditory stimulus would be occurring simultaneously with an attended or unattended visual stimulus. Thus, such a situation would not allow for the preset, early sensory enhancement of ERP responses to auditory stimuli that can occur when they are presented at the attended location as in the above-mentioned previous studies. In support of this explanation, in the multimodal attentional stream experiments described above, the interspersed unisensory visual and auditory stimuli need to be presented at the same location in space to reliably obtain supramodal attention effects on the irrelevant-modality unisensory stimuli; even a misalignment of  $3^\circ$  can eliminate spatial attention effects on auditory stimuli in vision-relevant conditions (37). Similarly, some recent intramodal auditory ERP studies have shown a rather steep gradient in the distribution of auditory spatial attention (falloff of  $\approx 60\%$  over  $3^\circ$ ) (41–42). Thus, the cross-modal

spread of attention in our paradigm (occurring despite the large 13.5° spatial separation between the visual and auditory stimuli) would seem to be due to the synchronicity of the visual and auditory stimuli. This synchronicity then resulted in the component sensory parts being grouped together into a multisensory object, such that if the visual component was attended, the attention would spread to encompass the cooccurring auditory input and enhance its processing.

Recently, an fMRI/ERP study investigated the dynamics of object-based attentional selection in the visual modality by using objects defined by color and motion (43). In line with our results, the authors reported that object-based attentional modulation arises later compared to the initial sensory processing latency. The size of this difference in latency ( $\approx 50$  ms) is somewhat smaller than the latency difference of the effect found in our study ( $\approx 120$  ms), which might indicate that the spread of attention takes longer across modalities than within one modality. Further experiments directly addressing this issue are needed to investigate any such cross-modal differences in the time needed to spread across object features.

The sustained attention-related difference found in the present study was maximal over frontal and fronto-central scalp regions (Fig. 3*b*). This scalp topography has similarity to the distribution of the auditory sensory N1 component at 100 ms that typically peaks over fronto-central scalp, reflecting a major contribution from a dipolar pair of auditory-cortex sources in the superior temporal plane that point upward and slightly forward (27). The distribution of the attentional spreading effect appears to be somewhat more frontal than a typical N1 component, thus also resembling the processing negativity component seen at earlier latencies (100–200 ms) in unimodal auditory attention experiments. The processing negativity is a prolonged, frontally distributed, negative wave that reflects additional activity elicited by attended auditory stimuli (relative to unattended stimuli) and that also appears to derive from sources in auditory cortex, with possibly some contribution from frontal cortex (25, 26, 40). Accordingly, the frontal/fronto-central distribution of the multisensory attentional spreading activity observed here might be due to a similar combination of auditory

cortex and frontal sources. The fMRI results in the present study confirm that the enhanced activity for the tones occurring synchronously with an attended visual event does indeed include enhanced activity in auditory cortex. We note, however, that the fMRI did not reveal any significant additional activity in frontal areas, and so it is not yet clear what additional frontal sources may be contributing to the ERP scalp distribution.

An auditory cortex contribution to the generation of the long-latency attentional effect observed here is in accord with the results for object-based attention in the visual modality. Schoenfeld *et al.* (43) report modulation of activity in ventral occipital cortical regions associated with the processing of color, in the case of object-based attentional spreading to color. Analogously, in the present experiment, the spread of attention from the visual to the auditory component of an audiovisual multisensory object is reflected by increased processing activity in auditory cortex.

In summary, we have shown that visual spatial attention can modulate the processing of irrelevant, spatially discrepant, auditory stimuli that are presented simultaneously with attended visual stimuli, with the latency of the effect indicating that this enhanced processing occurs relatively late in the processing stream. We interpret this attentional enhancement of the auditory processing as resulting from the grouping of the auditory and visual events, due to their temporal cooccurrence, into an audiovisual multisensory object. If the visual stimulus is attended, the cross-modal sensory grouping results in attention spreading beyond the attended spatial location and attended visual modality, such that it encompasses the spatially discrepant auditory stimulus, pulling it into the attentional spotlight and bestowing it with enhanced processing.

We thank T. Grent-'t-Jong, Chad Hazlett, and Roy Strowd for technical assistance and D. Talsma, H. Slagter, S. Katzner, and H.-J. Heinze for helpful comments. This research was supported by National Institute of Mental Health Grant R01-MH00415 and National Institute of Neurological Disorders and Stroke Grants P01-NS41328-Proj.2 and R01-NS051048 (to M.G.W.).

- Carrasco, M., Ling, S. & Read, S. (2004) *Nat. Neurosci.* **7**, 308–313.
- Posner, M. I. (1980) *Q. J. Exp. Psychol.* **32**, 3–25.
- Brefczynski, J. A. & De Yoe, E. A. (1999) *Nat. Neurosci.* **2**, 370–374.
- Hillyard, S. A. & Anillo-Vento, L. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 781–787.
- Moran, J. & Desimone, R. (1985) *Science* **229**, 782–784.
- Woldorff, M. G., Fox, P. T., Matzke, M., Lancaster, J. L., Veeraswamy, S., Zamarripa, F., Seabolt, M., Glass, T., Gao, J. H., Martin, C. C., *et al.* (1997) *Hum. Brain Mapp.* **5**, 280–286.
- Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. (1997) *J. Neurophysiol.* **77**, 24–42.
- Treue, S. & Trujillo, J. C. M. (1999) *Nature* **399**, 575–579.
- Saenz, M., Buracas, G. T. & Boynton, G. M. (2003) *Vis. Res.* **43**, 629–637.
- Egley, R., Driver, J. & Rafal, R. D. (1994) *J. Exp. Psychol. Gen.* **123**, 161–177.
- Blaser, E., Pylyshyn, Z. W. & Holcombe, A. O. (2000) *Nature* **408**, 196–199.
- O'Craven, K. M., Downing, P. E. & Kanwisher, N. (1999) *Nature* **401**, 584–587.
- Scholl, B. J. (2001) *Cognition* **80**, 1–46.
- Müller, N. G. & Kleinschmidt, A. (2003) *J. Neurosci.* **30**, 9812–9816.
- Stein, B. E. & Meredith, M. A. (1993) *The Merging of the Senses* (MIT Press, Cambridge, MA).
- Howard, I. P. & Templeton, W. B. (1966) *Human Spatial Orientation* (Wiley, New York).
- Driver, J. (1996) *Nature* **381**, 66–68.
- Bertelson, P. (1998) in *Advances in Psychological Science: Biological and Cognitive Aspects*, eds. Sabourin, M., Craik, F. I. M. & Roberts, M. (Psychological Press, Hove, U.K.), Vol. 1, pp. 419–439.
- Bertelson, P. & Radeau, M. (1981) *Percept. Psychophys.* **29**, 578–584.
- Pick, H. L., Jr., Warren, D. H. & Hay, J. C. (1969) *Percept. Psychophys.* **6**, 203–205.
- Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L. & Schirillo, J. A. (2003) *J. Cognit. Neurosci.* **15**, 20–29.
- Bertelson, P., Vroomen, J., de Gelder, B. & Driver, J. (2000) *Percept. Psychophys.* **62**, 321–332.
- Vroomen, J., Bertelson, P. & de Gelder, B. (2001) *Percept. Psychophys.* **63**, 651–659.
- Hillyard, S. A., Hink, R. F., Schwent, V. L. & Picton, T. W. (1973) *Science* **182**, 177–180.
- Woldorff, M. G. & Hillyard, S. A. (1991) *Electroencephalogr. Clin. Neurophysiol.* **79**, 170–191.
- Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D. & Bloom, F. E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8722–8726.
- Näätänen, R. & Picton, T. (1987) *Psychophysiology* **24**, 375–425.
- Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B. & Dale, A. M. (1998) *Neuron* **20**, 285–296.
- Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R. & Dale, A. M. (1998) *NeuroReport* **9**, 3735–3739.
- Talsma, D. & Woldorff, M. G. (2004) in *Event-Related Potentials: A Methods Handbook*, ed. Handy, T. (MIT Press, Cambridge, MA), pp. 115–148.
- Busse, L. & Woldorff, M. G. (2003) *NeuroImage* **18**, 856–864.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D. & Frackowiak, R. S. J. (1995) *Hum. Brain Mapp.* **2**, 189–201.
- Luck, S. J., Woodman, G. F. & Vogel, E. K. (2000) *Trends Cognit. Sci.* **4**, 432–440.
- Picton, T. W., Hillyard, S. A., Krausz, H. I. & Galambos, R. (1974) *Electroencephalogr. Clin. Neurophysiol.* **36**, 179–190.
- McDonald, J. J., Teder-Sälejärvi, W. A. & Hillyard, S. A. (2000) *Nature* **407**, 906–908.
- Stein, B. E., London, N., Wilkinson, L. K. & Price, D. D. (1996) *J. Cognit. Neurosci.* **8**, 497–506.
- Eimer, M. & Schröger, E. (1998) *Psychophysiology* **35**, 313–327.
- Talsma, D. & Kok, A. (2002) *Psychophysiology* **39**, 689–706.
- Teder-Sälejärvi, W. A., Münte, T. F., Sperlich, F. J. & Hillyard, S. A. (1999) *Brain Res. Cogn. Brain Res.* **8**, 327–343.
- Näätänen, R. (1982) *Psychol. Bull.* **92**, 605–640.
- Teder, W. & Näätänen, R. (1994) *NeuroReport* **5**, 709–711.
- Teder-Sälejärvi, W. A., Hillyard, S. A., Röder, B. & Neville, H. J. (1999) *Brain Res. Cogn. Brain Res.* **8**, 213–227.
- Schoenfeld, M. A., Tempelmann, C., Martinez, A., Hopf, J.-M., Sattler, C., Heinze, H.-J. & Hillyard, S. A. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 11806–11811.