

## Selective Attention and Audiovisual Integration: Is Attending to Both Modalities a Prerequisite for Early Integration?

Durk Talsma<sup>1,3</sup>, Tracy J. Doty<sup>1</sup> and Marty G. Woldorff<sup>1,2</sup>

<sup>1</sup>Center for Cognitive Neuroscience, Duke University, Durham, NC, USA, <sup>2</sup>Department of Psychiatry, Duke University, Durham, NC, USA and <sup>3</sup>Department of Cognitive Psychology, Vrije Universiteit, Amsterdam, The Netherlands

**Interactions between multisensory integration and attention were studied using a combined audiovisual streaming design and a rapid serial visual presentation paradigm. Event-related potentials (ERPs) following audiovisual objects (AV) were compared with the sum of the ERPs following auditory (A) and visual objects (V). Integration processes were expressed as the difference between these AV and (A + V) responses and were studied while attention was directed to one or both modalities or directed elsewhere. Results show that multisensory integration effects depend on the multisensory objects being fully attended—that is, when both the visual and auditory senses were attended. In this condition, a superadditive audiovisual integration effect was observed on the P50 component. When unattended, this effect was reversed; the P50 components of multisensory ERPs were smaller than the unisensory sum. Additionally, we found an enhanced late frontal negativity when subjects attended the visual component of a multisensory object. This effect, bearing a strong resemblance to the auditory processing negativity, appeared to reflect late attention-related processing that had spread to encompass the auditory component of the multisensory object. In conclusion, our results shed new light on how the brain processes multisensory auditory and visual information, including how attention modulates multisensory integration processes.**

**Keywords:** auditory, binding, electrophysiology, multisensory, visual

### Introduction

In order to focus on relevant information and ignore what is irrelevant, the human mind is equipped with a selection mechanism accomplished by the cognitive function of attention. Scientific studies of this mechanism have shown that attention is brought about in the brain by selectively increasing the sensitivity of perceptual brain areas that are responsive to the task-relevant stimulus feature, often combined with a simultaneous relative decrease in sensitivity of the perceptual brain areas that are responding to nontask-relevant stimulus features (e.g., Motter 1993; Hillyard and others 1995; Tootell and others 1998). A key role of attention is thus to serve the purpose of selectively enhancing perception (Hopfinger and Mangun 1998).

Although initially research had focused almost exclusively on the attentional processes that take place when selecting stimuli within a single sensory modality, an increasing number of contemporary studies are focusing on the dynamics of multisensory processing in selective attention. Among these studies, some using event-related potentials (ERPs) have established that selective attention is a mechanism that is not limited to a single sensory modality but can encompass or spread across multiple sensory systems (Eimer and Schröger 1998; Talsma and Kok 2001, 2002; Macaluso and others 2003; Busse and others 2005).

Adding to these findings, single-cell recordings in animals (Stein and Wallace 1996; Wallace and Stein 1997, 2001) as well as studies on the human electrophysiology (Fort and others 2002a, 2002b; Molholm and others 2002; Talsma and Woldorff 2005a) have established that information stemming from multiple senses is not likely to be processed in isolation but will tend to be integrated into a multisensory percept under various circumstances. Behavioral findings have shown that the near-simultaneous presentation of visual and auditory stimuli and their having common locations are 2 fundamental properties that facilitate the integration of audiovisual stimuli into a multisensory percept or multisensory object (Lewald and Guski 2003).

Several recent ERP studies have utilized the temporal proximity property of multisensory objects by adapting a method that was first developed in animal single-cell recordings (Stein and Meredith 1993) and that was then adapted for use in human ERP studies (Giard and Peronnet 1999). Using this approach, unisensory auditory (A), unisensory visual (V), and multisensory audiovisual (AV) objects are presented in random succession. Because the earliest ERP components reflect mainly sensory processing, multisensory integration processes can be studied by summing the unisensory auditory (A) and unisensory visual (V) ERPs and computing the difference between this summated (A + V) ERP and the ERP elicited by the simultaneous audiovisual (AV) stimuli. Thus, for the early ERP waves, integration can be expressed as a superadditive response, for example, when the ERP waves for the multisensory stimuli are larger than those of the summated responses of the visual and auditory unisensory stimuli (A + V). Using this method, it has been reported that the integration of auditory and visual stimulus properties into a multisensory object may take place relatively early on in the processing stream (Giard and Peronnet 1999; Molholm and others 2002). This finding suggests that integration is a process that occurs largely without conscious effort. In addition, many behavioral studies have provided evidence for the hypothesis that integrating visual and auditory stimuli serves the purpose of enhancing perceptual clarity (Stein and others 1996; Calvert and others 2000).

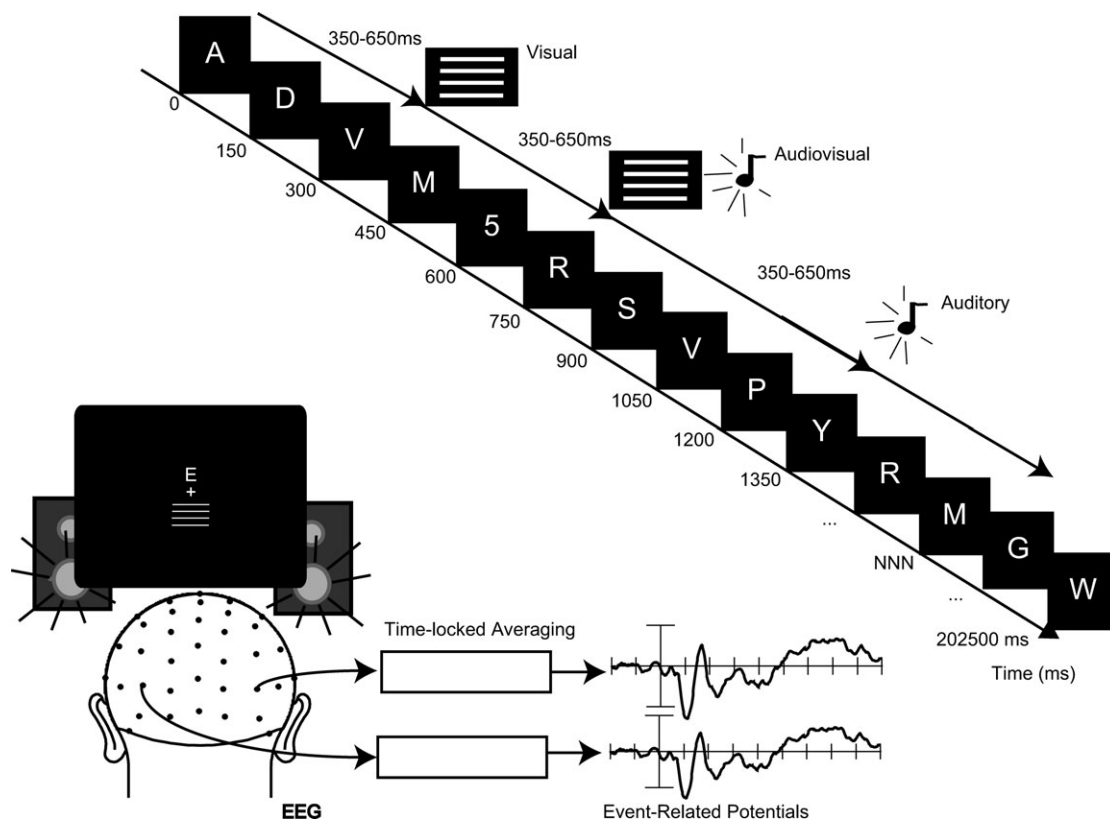
These results suggest that the communication between the visual and auditory brain areas is a highly effective and relatively automatic process (Foxy and others 2000). This suggestion has therefore given rise to a debate as to whether or not multisensory integration processes can be affected by attention. A number of behavioral studies have suggested that multisensory integration takes place at a preattentive stage and is not influenced by attention (Driver 1996; Bertelson and others 2000; Vroomen and others 2001a, 2001b).

In our previous work (Senkowski and others 2005; Talsma and Woldorff, 2005a), we predicted that multisensory integration

would nonetheless interact with attention because these processes both subserve the goal of enhancing perception. In those studies, interactions were indeed found in ERP waveforms by using the AV - (A + V) method. Attention was manipulated by presenting auditory, visual, and audiovisual stimuli randomly to 2 lateral spatial positions and instructing subjects to focus their attention at only one of these locations during a whole block of trials. When stimuli were presented at the attended location, we found that multisensory (AV) stimuli elicited larger ERP waveforms than the sum of the visual and auditory (A + V) parts elicited alone, whereas at the unattended location, the differences between the AV and A + V response were considerably smaller.

The onset of the multisensory integration ERP effects in the Talsma and Woldorff (2005a) study occurred at 100 ms post-stimulus, about 50 ms later in time than what had previously been reported in the literature (Giard and Peronnet 1999; Molholm and others 2002). One difference between our previous study and these other 2 ERP studies is that we presented stimuli peripherally, whereas the other studies presented both visual and auditory stimuli centrally (Giard and Peronnet 1999) or at a location particularly optimized to evoke early activity (Molholm and others 2002). Such a peripheral presentation in our previous study could have led to a situation where subjects were required to focus their attention strongly on the required location, which could in turn have led to 1) an enhancement of the observed attention effects on the multisensory integration process and 2) a slight delay of the integration process itself.

Additionally, in our prior studies, subjects were always required to focus their attention on both visual and auditory modalities, so that it remains unclear whether or not attending to just 1 of the 2 modalities will lead to integration processes or whether it is necessary to attend to both modalities. To resolve these issues, the current study sought to address the following 2 questions: 1) will central presentation lead to earlier effects of multisensory integration and 2) does the process of directing attention to one single sensory modality (i.e., attend auditory only or attend visual only) affect the process of integrating audiovisual stimulus features differently than attending to the visual and auditory modalities simultaneously? We sought to answer these questions by presenting visual-only, auditory-only, and audiovisual multisensory objects in central space, just below fixation. In addition, a rapid serial visual presentation (RSVP) letter stream was presented directly above fixation, which was used in one condition to direct attention away from the object stimuli (see Fig. 1). Subjects were given 4 different types of attentional instructions for the different runs. 1) In the attend RSVP condition, subjects were instructed to ignore both the visual and auditory objects. 2) In the attend auditory object condition, subjects were instructed to focus their attention on the auditory objects and the auditory part of the multisensory objects. 3) In the attend visual objects condition, subjects were instructed to attend to the visual objects and to the visual part of the multisensory objects. 4) In the attend audiovisual objects condition, subjects were instructed to attend to all objects (auditory, visual, and both modalities of the multisensory



**Figure 1.** General layout of the task. An RSVP stream of letters was presented just above fixation. Concurrently, at random (i.e., temporally jittered) moments in time, with an SOA of 350–650, visual, auditory, or audiovisual stimuli were presented in random order slightly below fixation. Stimulus presentation was such that each stimulus started completely randomly with respect to the onset of the letters in the RSVP stream.

objects). As noted above, in the attend RSVP condition, both visual and auditory object stimuli were considered to be unattended, even though attention was directed to the visual modality in this condition. These visual objects were considered to be attended because visual attention based on a nonspatial stimulus feature has been found to be generally contingent upon spatial selection (Van der Heijden 1992, 1993). To more specifically test this assumption, we analyzed the amplitudes of the early sensory components of the ERPs elicited by visual stimuli as a function of the different attention conditions. In particular, we examined the occipital P1 and N1 components, peaking at approximately 90–120 ms and 150–200 ms after stimulus onset, respectively. If the visual object stimuli were largely unattended in the attend RSVP condition, we expected that this would be reflected in lower P1 amplitudes (relative to the attend visual and audio-visual object conditions). In addition, we expected an occipital negative difference in the visual-object ERPs elicited in the attend (visual and audiovisual) object condition relative to the attend RSVP conditions, an effect known as the occipital selection negativity. This effect typically occurs between about 200 and 300 ms after stimulus onset (Kenemans and others 1993, 2002; Talsma and Kok 2001). For auditory attention effects, nonspatial attention is typically expressed in a series of processing negativities. Thus, we expected to observe such a series of effects when subjects were attending the auditory stimulus objects.

Because visual objects were presented from a midline location, relatively close to the fovea, with the apparent location of the auditory stimulus being matched to that of the visual one, we expected that multisensory integration ERP effects might occur earlier in time than what we found in our previous study (Talsma and Woldorff 2005a). Based on our previous findings, we expected that attention could also interact with these early effects of integration, in particular that attending to both modalities would lead to an increase in audiovisual integration processes.

How the multisensory integration processes would be affected by attention when only one modality is attended also remained to be elucidated, because to our knowledge, no studies have addressed this question yet. Therefore, this question remains somewhat exploratory. It has been found, however, that auditory stimuli are known to capture attention easily (Schröger and others 2000) and also that the processing of auditory stimulus features occurs generally faster than that of visual stimuli (Woldorff and others 1991, 1993). Thus, on the basis of these differences in processing time, we would also predict that attending to the visual modality would affect the multisensory integration processes differently than attending to the auditory modality.

## Methods

### Participants

Twenty healthy volunteers participated in the experiment (aged 18–25 years, mean 19.2; 13 males and 7 females). All participants had normal or corrected-to-normal vision and normal hearing capabilities. Participants were recruited through local advertisements at the campus of Duke University and were paid \$10/h for their participation or received credit for their participation as part of a requirement for an introductory psychology class at Duke University. All participants gave written informed consent for their participation. One participant was excluded from the analyses due to poor data quality.

### Stimuli and Task

The task described here consisted of the combined presentation of an RSVP letter stream along with auditory and visual objects (see Fig. 1). The letter stream was presented directly above fixation (1 degree). Letters were sequentially presented, being randomly replaced every 150 ms. This random replacement was restricted in such a way that a letter was always replaced with a different letter. Every 1–10 s, randomly, a digit was presented instead of a letter, which served as the target stimulus when subjects were attending the letter stream (see below).

The auditory and visual objects were presented either separately (unimodal presentation) or simultaneously (multimodal presentation). Unimodal visual stimuli consisted of white horizontal square wave gratings subtending a 5-degree visual angle presented against a black background. These visual stimuli were presented directly below (~3.5 degree) the central fixation point, each with a duration of 105 ms.

Unimodal auditory stimuli consisted of a 1600-Hz tone pip (duration of 105 ms, linear rise and fall times of 10 ms), presented at –65 dB SPL(a). These stimuli were presented through 2 speakers placed slightly lateral to and behind the monitor, such that the speakers were hidden from the subject's view. Auditory stimuli were presented simultaneously from the two speakers, such that the subjective location of the auditory stimuli matched the location of the visual objects (Eimer and Schröger 1998). Multisensory stimuli consisted of a combination of both auditory and visual features. Presenting both the visual and auditory stimuli simultaneously created the subjective impression of a single multisensory audiovisual object.

Subjects were given 4 types of attentional instructions but in all cases were instructed to keep their eyes focused on the fixation cross and direct their attention covertly to a designated subset of the presented objects—1) Attend RSVP: Subjects were instructed to focus their attention on the RSVP letter stream and detect and respond to the target digits. 2) Attend audiovisual: Subjects were instructed to attend to all the visual, auditory, and audiovisual objects and to detect occasional targets (20% of all stimuli) in both the visual and auditory modalities. Target stimuli were highly similar to standards but contained a transient dip in intensity halfway through the duration of the stimulus, which caused the subjective impression that the stimulus appeared to flicker (visual target) or to stutter (auditory target). The degree of intensity reduction was determined for each subject individually during a training session prior to the experiment (Senkowski and others 2005; Talsma and Woldorff 2005a). Multisensory targets always contained the midstimulus intensity decrease in both the visual and auditory modalities. 3) Attend visual only: Subjects were instructed to attend to the visual objects and to only the visual components of the multisensory objects in order to detect visual targets among these. Targets were the same stimuli as described in the attend audiovisual condition above. 4) Attend auditory only: Subjects were instructed to attend to auditory objects and to only auditory components of the multisensory stimuli in order to detect auditory targets among these. In all conditions, subjects were required to report the targets by making a speeded button-press response on a game pad controller device. To summarize, in addition to the RSVP letter stream, 6 different stimulus categories (trial types) were used in the present experiment consisting of the combination of stimulus modality (3 levels: unimodal visual, unimodal auditory, or multimodal audiovisual) and stimulus type (2 levels: targets or standards).

A computer generated each new first-order counterbalanced, randomized object-stimulus order and randomized RSVP letter sequence for each subject. The stimulus onset asynchrony (SOA) of the objects varied randomly between 350 and 650 ms (mean SOA 500 ms), which was completed uncorrelated with respect to the onset of the letters in the RSVP stream. For each condition (attend RSVP/attend multisensory/attend visual only/attend auditory only), 200 visual, 200 auditory, and 200 multisensory stimuli were presented; of these 200 stimuli, 160 stimuli were standards and the remaining 40 stimuli in each category were targets.

### Procedure

To familiarize participants with the stimulus material, they were first given the discrimination task that determined the individual subject's target discrimination thresholds (Talsma and Woldorff 2005a). After this session was completed, the electrocaps were put in place and

participants were seated and given the task-specific instructions, along with a number of practice blocks. Participants continued training until the experimenter was convinced that the participants understood the task. To avoid movement artifacts, participants were further instructed to minimize blinking and making body movements and to fixate on a centrally presented fixation dot. Prior to each run, participants were instructed which stimulus to attend to, and after the run was completed, they were given feedback about their performance. Participants were allowed to take short breaks between runs.

### Apparatus

Stimulus presentation was controlled by a personal computer running the "Presentation" software package (Neurobehavioral Systems, Inc., Albany, CA). ERPs were recorded from 64 equally spaced tin electrodes, mounted in a custom-designed elastic cap (Electro-Cap International, Inc., Eaton, OH) and referenced to the right mastoid during recording. In the remainder of this paper, electrodes will be referred to by their approximate position relative to the standard international 10-10 system, with a suffix providing additional localization information. More specifically, an electrode with a position slightly inferior to the standard location (within 1.0-1.5 cm) is indicated using a suffix of "i". Similarly, a suffix of "a" or "p" indicates that this electrode was placed slightly anterior or posterior to the standard location.

Electrode impedances were kept below 2 kΩ for the mastoids and ground, 10 kΩ for the eye electrodes, and 5 kΩ for the remaining electrodes. Horizontal eye movements were monitored by 2 electrodes at the outer canthi of the eyes. Vertical eye movements and eye blinks were detected by electrodes placed below the orbital ridge of both eyes, which were referenced to 2 electrodes directly located above the eyes. During recording, eye movements were also monitored using a closed circuit video monitoring system. Electroencephalography (EEG) was recorded using a Neuroscan (SynAmps) acquisition system using a band-pass filter of 0.01-100 Hz and a gain of 1000. Raw signals were continuously digitized with a sampling rate of 500 Hz and digitally stored for off-line analysis. Recordings took place in a sound-attenuated, dimly lit, electrically shielded room.

### Data Analysis

#### Behavioral Reaction Times

Reaction times (RTs) for correct detections of targets, hit rates (HRs), and false alarm rates were computed separately for the different conditions. These measures were subjected to an analysis of variance (ANOVA) with stimulus type (3 levels: multisensory, unisensory visual, or unisensory auditory) as a within-subject factor. In addition, mean RTs and HRs to the RSVP targets in the attend RSVP were calculated.

#### ERP Analysis

Artifact rejection was performed off-line by discarding epochs of the EEG that were contaminated by eye movements, eye blinks, excessive muscle-related potentials, drifts, or amplifier blocking, according to the methods described in Talsma and Woldorff (2005b). Approximately one-third of the trials were rejected due to artifacts leaving about 100-110 artifact-free trials for inclusion in each single-subject average. Averages were calculated for the different stimulus types from 1000 ms before to 1200 ms after stimulus onset. The averages were digitally filtered with a noncausal, running-average filter of 9 points, which strongly reduced signal frequencies at and above 56 Hz at our sample rate of 500 Hz. After averaging, all channels were rereferenced to the algebraic average of the 2 mastoid electrodes. The adjacent response (ADJAR) procedure (Woldorff 1993) was used to estimate and remove distortions of the ERP waves due to overlapping trial sequences. ADJAR is an iterative process that with increasing iteration converges to optimal overlap estimates. For each subject, it took approximately 145-155 iterations for the overlap estimates to fully converge.

#### Statistical Analyses

Two types of analyses were conducted. First, the effects of selective attention were established separately for the visual, auditory, and multisensory stimuli. Second, the effects of multisensory integration and interactions between attention and multisensory integration were

determined. Statistical analyses of the early ERP components, such as P50, P1, and N1, were conducted as follows: First, peak latencies were computed separately for each subject/condition individually, using a subset of electrodes that was established on the basis of prior visual selection. The minimum and maximum latencies were individually adjusted to ensure that each individual component's peak was enclosed in the search window. Then, mean amplitudes of these components were computed using a small window surrounding the peak maximum. Mean amplitudes were chosen instead of peak amplitudes because the (AV - [A + V]) transformations resulted in ERP waveforms that were composed of different trial numbers, which might otherwise have resulted in biased peak-amplitude measures (Handy 2005). The width of the window was individually determined for each peak component separately and will be reported in Results where appropriate.

Longer latency ERP waveforms were tested by computing mean amplitudes using consecutive windows of 20 ms each. These measures were also computed on the basis of a selection of electrodes where visual inspection of the waveforms and scalp topography plots had shown these differences to be most distinguishable. For all tests, within-subject ANOVA was used to determine the significance of differences between conditions. Greenhouse-Geisser correction was applied for tests involving factors with more than 2 levels. The specific factorial design is given in Results where appropriate.

## Results

### Behavioral Data

#### Reaction Times

Table 1 presents the mean RTs for each stimulus type. For each attention condition (attend visual, attend auditory, and attend audiovisual), the RTs to visual-only and to auditory-only objects were compared with the RTs to audiovisual objects, using pairwise *t*-tests. In the attend auditory condition, no significant RT difference between auditory and audiovisual objects could be found ( $T_{18} = 1.77, P > 0.1$ ). In the attend visual condition, RTs to visual-only objects were significantly faster than RTs to audiovisual objects ( $T_{18} = 4.57, P < 0.001$ ). In the attend audiovisual condition, no significant RT difference could be found between visual and multisensory stimuli ( $T_{18} = 0.44, P > 0.6$ ), but RTs were significantly slower. Reactions were significantly slower to auditory than to audiovisual stimuli, however ( $T_{18} = 4.46, P < 0.0005$ ). Finally, whereas in the 2 unisensory conditions (attend auditory and attend visual) the RTs to the 2 unisensory stimuli did not differ significantly from each other ( $T_{18} = 0.07, P > 0.9$ ), in the attend audiovisual condition, the RT to the auditory stimulus was significantly slower than that to the visual stimulus ( $T_{18} = 2.43, P < 0.02$ ). Mean RT to the target digits in the RSVP condition was 598 ms.

#### Accuracy

HRs are given in Table 2. Accuracy did not differ between multisensory targets and auditory targets in the attend auditory condition ( $T_{18} = 1.2, P > 0.2$ ). In the attend visual condition, responses to visual-only stimuli were more accurate than those

**Table 1**  
Mean response times to the target stimulus objects

	Stimulus		
	Auditory	Visual	Audiovisual
Attend auditory	508 (78)		495 (66)
Attend visual		509 (54)	571 (58)
Attend audiovisual	579 (79)	533 (57)	525 (73)

Note: All times in milliseconds. Standard deviations are given in parentheses.

**Table 2**

Percentage correctly reported of target stimulus objects

	Stimulus		
	Auditory	Visual	Audiovisual
Attend auditory	84 (18)		86 (15)
Attend visual		79 (17)	61 (21)
Attend audiovisual	71 (22)	70 (18)	81 (15)

Note: Standard deviations are given in parentheses.

**Table 3**

N1 amplitudes

	AV	A + V
Attend RSVP	-6.40 (2.7)	-6.18 (3.0)
Attend auditory	-6.92 (2.4)	-7.89 (3.1)
Attend visual	-7.36 (4.2)	-7.02 (4.3)
Attend audiovisual	-7.46 (3.5)	-5.07 (3.2)***

Note: All amplitudes are in microvolts. Standard deviations are given in parentheses (\*\*\*)significant;  $P < 0.0005$ .

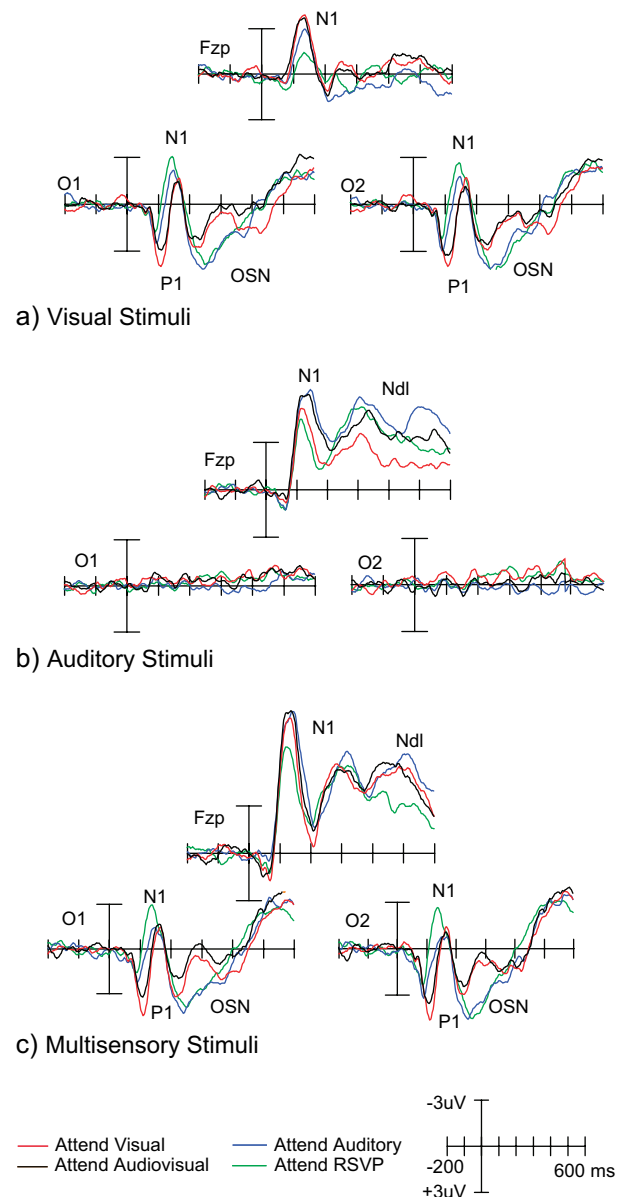
to multisensory stimuli ( $T_{18} = 4.30$ ,  $P < 0.0004$ ). In the attend audiovisual condition, responses to audiovisual stimuli were slightly more accurate than those to either visual ( $T_{18} = 2.21$ ,  $P < 0.05$ ) or auditory stimuli ( $T_{18} = 3.25$ ,  $P < 0.005$ ) alone. Finally, response accuracy to unisensory visual and auditory stimuli did not differ in either the unisensory attention conditions ( $T_{18} = 0.90$ ,  $P > 0.3$ ) or the attend audiovisual condition ( $T_{18} = 0.24$ ,  $P > 0.81$ ). Mean HR to the RSVP target digits in the attend RSVP condition was 76%.

### Event-Related Potentials

#### Selective Attention Effects

**Visual stimuli.** ERPs elicited by visual objects consisted of occipital P1, N1, and P2 components as well as an anteriorly recorded N1 component. Attention effects on these visual ERPs consisted mainly of an enhancement of the occipital P1 and of the anterior N1 components, followed by a spatially broader selection negativity over the occipital areas, between about 200 and 300 ms after stimulus onset (see Fig. 2a). These effects were statistically analyzed by collapsing the ERPs to the visual objects in the attend visual and attend audiovisual conditions into a single “attended” category (after determining that there was no significant difference between these 2 conditions) and collapsing the ERPs to the visual objects in the attend auditory and attend RSVP conditions into a single “unattended” category (also after determining that there was no significant difference between these 2 conditions). The P1 effect was statistically tested by determining the peak latency of each peak and computing the mean amplitude of a 50-ms window around the peak, on channels O1 and O2. These amplitudes were then subjected to ANOVA, using attention (attended or unattended) and channel (left or right hemisphere) as within-subject factors. The P1 was significantly larger for attended than for unattended visual objects ( $F_{1,18} < 6.6$ ,  $P < 0.02$ ). In addition, it peaked significantly later in the attended conditions (101 ms) than in the unattended conditions (90 ms) ( $F_{1,18} = 22.41$ ,  $P < 0.0002$ ).

In contrast to the P1 amplitude, Figure 2 suggested that the N1 amplitude, relative to baseline, was actually smaller in the



**Figure 2.** Summary of attention effects. (a) Attention effects on the unisensory visual stimuli. At the occipital channels, a clear enhancement of the P1 component could be seen for the attend visual condition and to a somewhat lesser degree also for the attend audiovisual condition. Although there appeared to be an N1 effect that followed an inverse pattern to that of the P1 effect, peak-to-peak amplitude testing showed that this effect on the unisensory visual stimuli was not significant, and the N1 shifts appeared therefore to be largely determined by the amplitude of the preceding time-extended P1 effects. Attended stimuli (attend visual and attend audiovisual) were further characterized by an occipital selection negativity between 220 and 350 ms. (b) Attention effects on the auditory stimuli consisted of an enhancement of the N1 component for the attend auditory and attend audiovisual conditions. (c) Attention effects on the multisensory stimuli. These effects consisted largely of a combination of visual and auditory effects.

attended conditions than in the unattended conditions. Statistical testing showed that this effect did not quite reach significance ( $F_{1,18} = 3.76$ ,  $P = 0.06$ ). Figure 2 also suggested, however, that the N1 component was being partially overlapped by a time-extended positivity following the P1, particularly in the attended conditions. Subsequent analysis showed, indeed, that the N1 correlated inversely with the amplitude of the preceding P1 component. That is, the (negative) baseline-to-peak

amplitude of the N1 component was smaller when the (positive) baseline-to-peak amplitude of the P1 was larger (Pearson's  $r = 0.37$ ,  $T_{150} = 4.8$ ,  $P < 0.00001$ ), fitting with the idea that a time-extended P1 effect may have been overlapping onto the N1 wave and shifting it down. We therefore reexamined the amplitude of the N1 relative to the preceding P1 using a peak-to-peak amplitude measure. More specifically, the latency of the maximum negative amplitude between 100 and 200 ms was determined. Then the mean amplitude in a 50-ms window surrounding this peak was determined, relative to the mean amplitude of the preceding P1. These values were submitted to a similar ANOVA as described above. When this approach was used for measuring N1 amplitude, no significant effects on the N1 could be found here.

The anterior N1 was tested at electrode Fz, also by determining the latency of the peak and by determining the mean amplitude of a small 50 ms around the peak. The anterior N1 peaked at 138 ms after stimulus onset for the attended conditions and at 127 ms after stimulus onset in the unattended conditions ( $F_{1,18} = 4.62$ ,  $P < 0.05$ ). In addition, the anterior N1 was larger in amplitude in the attended conditions than in the unattended conditions ( $F_{1,18} = 11.9$ ,  $P < 0.005$ ).

Because the selection negativity is a relatively slow endogenous wave, the significance of this effect was tested by computing mean voltages in consecutive windows of 20 ms each, also at electrodes O1 and O2. This effect became significant starting at 220 ms after stimulus onset and lasting until 300 ms ( $4.63 < F_{1,19}$  values  $< 13.38$ ,  $0.002 < P$  values  $< 0.05$ ).

**Auditory stimuli.** Attended auditory stimuli (i.e., in the attend auditory and attend audiovisual conditions) showed an enhancement of the frontocentral N1 (see Fig. 2b). This effect was tested using a similar approach as described for the visual stimuli (i.e., by collapsing auditory ERPs from the attend auditory and attend audiovisual conditions [attended] as well as collapsing auditory ERPs from the attend visual and attend RSVP conditions [unattended]), determining the latency of the peak and testing a 50-ms window surrounding the peak, performed at electrode Fz. This analysis showed that the frontocentral N1 peaked earlier in the unattended conditions than in the attended conditions (115 vs. 132 ms,  $F_{1,18} = 30.9$ ,  $P < 0.00001$ ). Peak amplitude was significantly larger in the attended conditions than in the unattended conditions ( $F_{1,18} = 15.4$ ,  $P < 0.001$ ). Possible effects of attention at longer latencies, as suggested by Figure 2b, were investigated by testing consecutive 20-ms time windows, but these tests failed to reach significance.

A final test on the auditory stimuli was conducted to determine whether intermodal attention to the auditory stimuli would evoke a late processing negativity. This was done by computing mean voltages in consecutive 20-ms windows on electrode Fz and subjecting these to ANOVA with attention (attend visual vs. attend auditory) as a within-subject factor. In this analysis, significant effects were found between 280 and 400 ms ( $F_{1,18}$  values = 5.12–9.21,  $P$  values  $< 0.05$ –0.01) and between 440 and 580 ms ( $F_{1,18}$  values = 4.85–8.12,  $P$  values  $< 0.05$ –0.01).

**Audiovisual stimuli.** ERPs elicited by audiovisual stimuli consisted largely of the combined activity elicited by visual and auditory stimuli alone. More specifically, these ERPs consisted largely of occipital P1 and N1 waves characteristic of visual ERP activity, in combination with a more anteriorly distributed N1 that is characteristic of an auditory ERP. These

effects were tested using the same tests as those conducted for the visual and auditory stimuli. However, because different components of the multisensory stimuli were attended in the 4 attention conditions, these effects were tested using 4 levels for the factor attention (attend visual, auditory, multisensory, or RSVP). This analysis showed a main effect of attention on the P1 ( $F_{3,54} = 4.28$ ,  $P < 0.025$ ). P1 amplitudes did not differ significantly between the attend RSVP and attend auditory conditions ( $F_{1,18} = 3.76$ ,  $P < 0.07$ ) but did between attend RSVP and attend visual ( $F_{1,18} = 11.5$ ,  $P < 0.005$ ), as well as between the attend RSVP and attend audiovisual conditions ( $F_{1,18} = 6.22$ ,  $P < 0.05$ ).

The selection negativity for the multisensory stimuli became significant between 240 and 300 ms after stimulus onset ( $5.50 < F_{3,57}$  values  $< 10.98$ ,  $0.002 < P$  values  $< 0.001$ ). The effects during this time window were reminiscent of a selection negativity similar to what we observed for the visual-only stimuli. In the 2 conditions in which the visual component of the multisensory stimulus was attended (i.e., in the attend visual and attend audiovisual conditions), the ERP waveforms were negatively displaced relative to the 2 other conditions (attend auditory and attend RSVP) in which the visual component of this stimulus was unattended (see Fig. 2c). Significant differences were found neither between the attend visual and attend audiovisual conditions nor between the attend auditory and attend RSVP conditions.

The baseline-to-peak amplitude test on the N1 revealed that a main effect of attention ( $F_{3,54} = 2.89$ ,  $P = 0.06$ ) was near significance. Similar to the unisensory visual stimuli, however, the amplitude of the N1 appeared to be shifted by a partially overlapping, time-extended P1 effect (see Fig. 2). Therefore, the amplitude of the N1 was reanalyzed using a peak-to-peak test (P1–N1). Using this approach, we could find no significant effects of attention ( $F_{3,54} = 2.36$ ,  $P > 0.1$ ). The occipital N1 component did peak at slightly different times in the 4 conditions: 136 ms (attend RSVP), 143 ms (attend auditory), 153 ms (attend visual), and 148 ms (attend audiovisual) ( $F_{3,54} = 4.18$ ,  $P < 0.05$ ).

The frontal N1 component was also significantly modulated by attention ( $F_{3,54} = 6.04$ ,  $P < 0.005$ ). Interestingly, during this time window, the amplitude of the N1 component for the multisensory objects was reduced when subjects were attending the RSVP letters, whereas the amplitude of this component was similar to when subjects were attending to any aspect (visual, auditory, or both visual and auditory) of the multisensory stimulus. No later significant effects of attention were found over frontal areas.

#### *Early P50 Modulations of Integration and Attention*

**Analyses.** The interactions between attention and audiovisual integration were determined by measuring the P50 amplitudes at electrodes FCz, Cz, and Pz, using the peak-picking method described earlier. Both amplitude and latency measures were submitted to ANOVA with the within-subject factors stimulus type (AV vs. [A + V]) and attention (attend RSVP, attend auditory, attend visual, and attend audiovisual).

**Timing.** The P50 latency differed significantly across the 4 attention conditions ( $F_{3,54} = 3.45$ ,  $P < 0.025$ ). Mean P50 latency was 59 ms in the attend RSVP condition, 60 ms in the attend auditory condition, 62 ms in the attend visual condition, and 55 ms in the attend audiovisual condition. In addition, we found an

interaction between stimulus type and attention ( $F_{3,54} = 3.40$ ,  $P < 0.05$ ). Post hoc testing showed that this interaction could be mainly explained by a significant latency difference between AV and A + V stimuli in the attend audiovisual condition ( $F_{1,18} = 5.15$ ,  $P < 0.03$ ).

**ERP waveforms.** As suggested in Figure 3, the amplitudes of the early P50 components of unisensory and multisensory stimuli depended significantly on attention, which was expressed in a significant interaction between the factors stimulus type and attention ( $F_{3,54} = 3.11$ ,  $P < 0.05$ ). In the attend audiovisual condition, the P50 amplitude elicited by multisensory stimuli was significantly larger than the combined activity elicited by the sum of the unisensory auditory and visual stimuli ( $F_{1,18} = 5.47$ ,  $P < 0.05$ ). In contrast, in the RSVP condition, when these objects were unattended, this pattern was reversed, with the P50 amplitudes being significantly smaller in the multisensory ERPs than in the summated unisensory ERPs ( $F_{1,18} = 4.21$ ,  $P < 0.05$ ). No significant effects on the P50 were found between the attend auditory ( $F_{1,18} = 1.35$ ,  $P > 0.2$ ) and attend visual conditions ( $F_{1,18} = 0.4$ ,  $P > 0.5$ ). In summary, multisensory attention interactions on P50 amplitude were found only when participants were either fully attending both modalities simultaneously or not attending the objects, with the direction of the modulation being reversed between the 2 conditions.

**Scalp topographies.** To further assess whether the observed modulations were manipulations of the P50, scalp topographies of this effect were analyzed for the attend RSVP and attend audiovisual conditions separately (see Fig. 4). These analyses were conducted by using the topography-normalized voltages (McCarthy and Wood 1985) from a subset of frontocentral channels (F7a, F3i, C3a, C5a, 3a, F3s, FC1, C1a, AFz, Fz, FCz, Cz, F4a, F4s, FC2, C2a, F8a, C4a, C6a) as input for a within-subjects

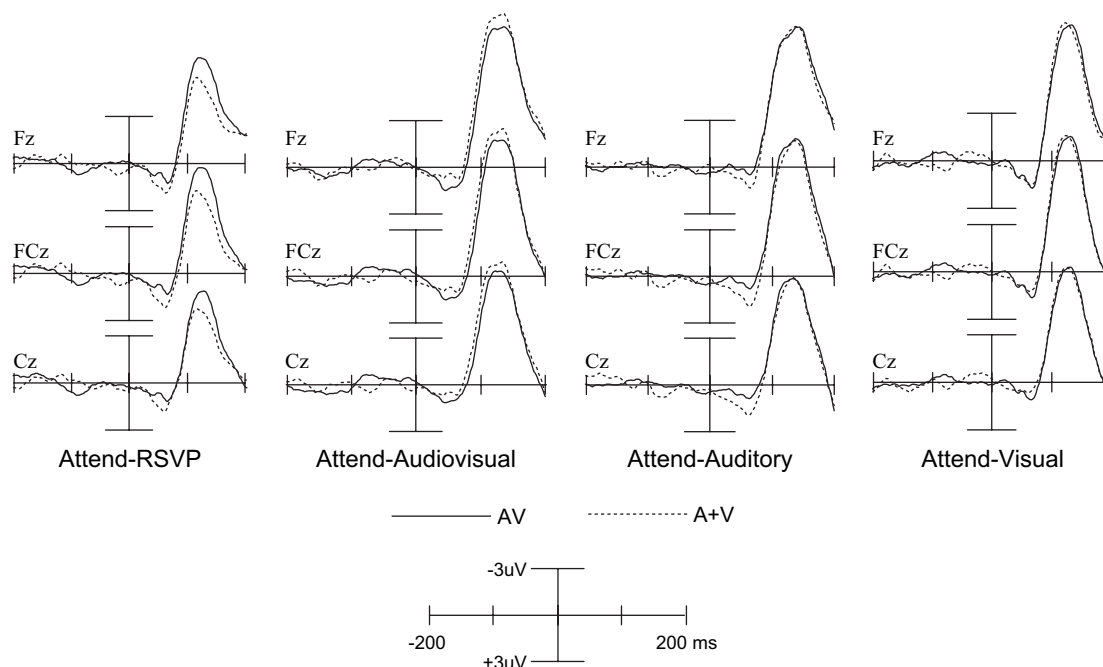
ANOVA. The within-subject factors of this ANOVA were multisensory effect (AV - [A + V] difference wave) versus the unisensory auditory P50 in that condition, laterality (5 levels: left lateral, left medial, midline, right medial, and right lateral), and anterior-posterior position (4 levels: frontal, frontocentral, central, and posterior). In this analysis, significant interactions of the factor multisensory effect with any of the other 2 factors are of particular interest (cf. Talsma and Kok 2001) as they signify a difference in topography between conditions and therefore a difference in the underlying neural configuration.

In the attend RSVP condition, these analyses indicated that the scalp topography of the ([A + V] - AV) multisensory P50 effect did not differ from that of the unisensory auditory P50 ( $P > 0.15$ , for all interactions). Similarly, the P50 topography of the (AV - [A + V]) multisensory effect did not differ from that of the unisensory auditory P50 in the attend audiovisual condition ( $P > 0.31$ , for all interactions). Based on these results, there is no basis to conclude that the P50 effects reflected in the (AV - [A + V]) contrasts are originating from brain areas other than those generating the P50 component.

#### N1 Effects

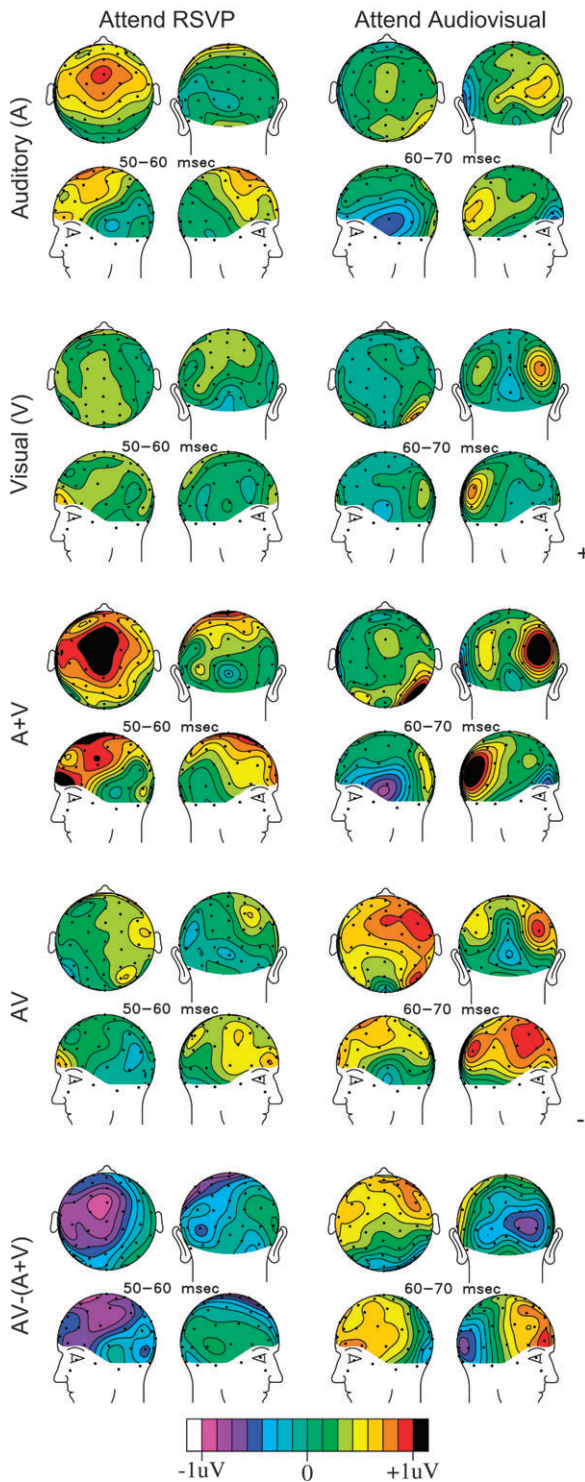
In addition to the early P50 modulations described above, (AV - [A + V]) interactions could be observed on the frontocentral N1 component that followed the P50. The N1 amplitude was also determined using the peak-finding procedure, using a time window of 100–200 ms after stimulus onset, on electrodes FCz, Cz, and Pz.

An overall ANOVA using the factors attention (4 levels: attend RSVP, attend auditory, attend visual, or attend audiovisual), stimulus type (2 levels: AV or [A + V]), and electrode (3 levels: FCz, Cz, and Pz) revealed a significant interaction between attention and stimulus type ( $F_{3,54} = 4.37$ ,  $P < 0.02$ ). Post hoc comparisons revealed that this effect was largely driven by



**Figure 3.** Multisensory integration by attention interactions on the frontocentral P50 and N1 components. P50 components for the multisensory stimuli were significantly larger in the attend audiovisual condition than for the summed unisensory responses, but this effect was reversed for the attend RSVP condition. Although the figure suggests an effect in the attend auditory condition, this effect was not statistically significant.

## P50 Audiovisual Integration Effects



**Figure 4.** Scalp topographies of the P50 components in the attend RSVP and attend audiovisual conditions.

a significant difference in N1 amplitude between unisensory and multisensory stimuli in the attend audiovisual condition ( $T_{18} = 4.3$ ,  $P < 0.0005$ ). In this condition, the N1 was significantly larger for multisensory stimuli than for unisensory stimuli (see table 3). No other significant N1 amplitude differences between multisensory and unisensory stimuli were found.

## Late Processing Negativity Effects

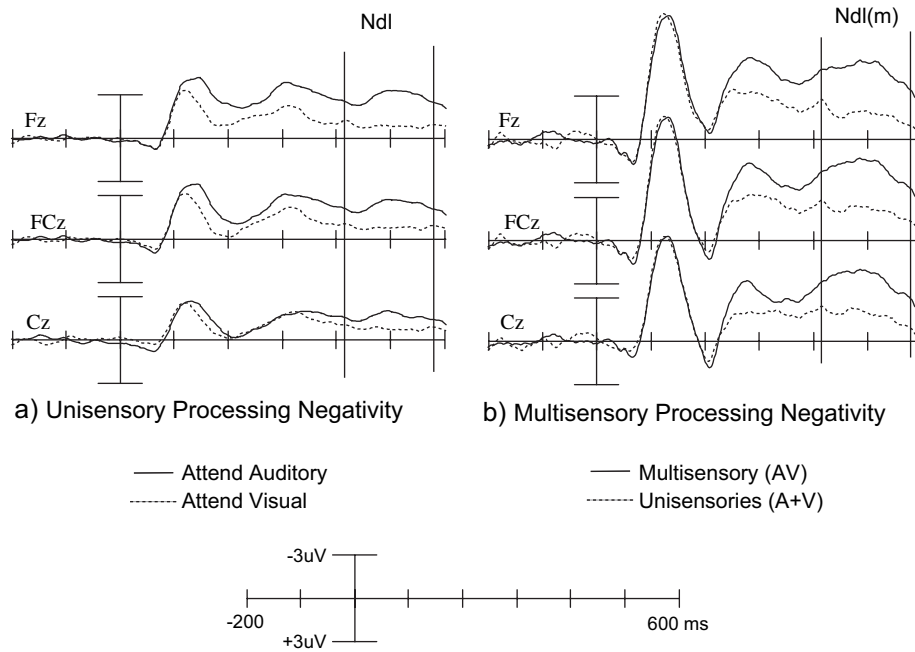
**Multisensory processing negativity.** Starting around 250 ms after stimulus onset, in the attend visual condition, the frontal activity elicited by multisensory stimuli showed a second, slower, negative deflection, which was much larger than that for the combined ERP activity of unisensory auditory and visual stimuli (Fig. 5b). The time course of this slow-wave activity was tested by analyzing consecutive 20-ms mean amplitudes of the  $(AV - [A + V])$  difference wave in the attend visual versus the attend auditory conditions. This effect was also expressed in a significant effect of stimulus type between 420 and 600 ms after stimulus onset (tests for all these 20-ms windows:  $5.74 < F_{1,19}$  values  $< 14.29$ ,  $0.001 < P$  values  $< 0.05$ ). As will be discussed later, this frontocentral multisensory ERP activity, elicited by visual attention, is reminiscent of the auditory processing negativities that are often elicited by attended versus unattended auditory stimuli and that were seen in this study.

**Scalp topographies.** The similarity in scalp topography between the unisensory and multisensory processing negativities was tested by computing a mean amplitude of both the unisensory and multisensory processing negativities, across the 420- to 580-ms time window, where both were significant using the same subset of channels as used for the P50 topography (F7a, F3i, C3a, C5a, 3a, F3s, FC1, C1a, AFz, Fz, FCz, Cz, F4a, F4s, FC2, C2a, F8a, C4a, C6a). Again, in this analysis, the interaction between channel and condition is of particular importance because it signifies differences in topographic distribution between these 2 effects. Confirming the observation in Figure 5, this interaction was indeed not significant ( $F < 1$ ). Based on this analysis, there is no basis to conclude that the multisensory processing negativity is generated by other neural structures than those generating the auditory processing negativity (see Figure 6).

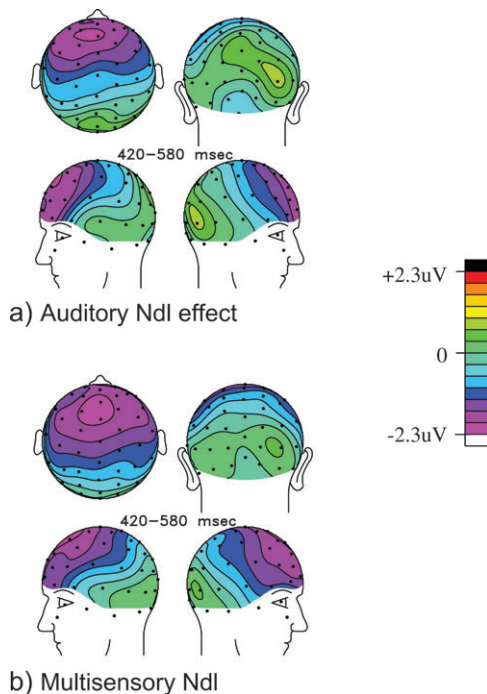
## Discussion

The present study investigated the effects of specifically attending to visual and/or auditory modalities on the integration of multisensory objects presented in central space. In this paper, we report 3 main new findings. First, when the auditory, visual, and audiovisual objects were attended, the P50 to the audiovisual stimuli was larger than the sum of the P50 activity for the auditory and visual stimuli, whereas when these stimuli were unattended, this audiovisual integration effect was reversed (i.e., smaller for the audiovisual response than for the sum of the unisensory ones). Second, although early physiological and behavioral effects of multisensory integration could be observed in this study, both of these effects appeared to critically depend on the subject attending to both modalities simultaneously. Third, audiovisual integration processes appeared to associate the visual and auditory stimulus components with each other, even when only the visual component was relevant. This was reflected in the spread of enhanced processing from the visual to the auditory modality. The latter effect was suggested by the presence of a late frontal negativity for the audiovisual stimuli in the attend visual condition that bore a strong similarity to the auditory late processing. These 3 results will be discussed in detail below.





**Figure 5.** ERP waveforms for the unisensory and multisensory late processing negativities (Ndl). (a) Unisensory processing negativity obtained by contrasting ERPs to auditory stimuli in the attend auditory and attend visual conditions. (b) Multisensory processing negativity obtained by contrasting AV and [A + V] stimuli in the attend visual conditions. See main text for details.



**Figure 6.** Scalp topographies of the unisensory and multisensory processing negativities depicted in Figure 5.

### Reversed Integration Effects for Attended versus Unattended Stimuli

One main finding of the present study was that the earliest effects of multisensory integration appeared to invert as a function of attention. A behavioral advantage of processing audiovisual stimuli was found only when subjects were attend-

ing to both modalities and was expressed in higher response accuracies to multisensory stimuli compared with unisensory stimuli. Whereas in the attend audiovisual condition the earliest positive polarity components elicited by multisensory stimuli were larger than those found in the summated ERP waveforms to the unisensory stimuli, in the attend RSVP condition (i.e., attending away from the multisensory objects) the opposite effect was found. Here these components were actually significantly smaller for multisensory than for the summed unisensory. Using scalp topography analyses, we could find no evidence to suggest that these early enhancement and depression effects were generated by brain structures other than those generating the P50. We therefore suggest that these effects are modulations of the P50 amplitude.

Single-cell animal studies have reported 2 principal multisensory interaction response patterns in the brain, particularly of superior colliculus neurons that have been shown to be under the influence of cortical control (Jiang and Stein 2003). In addition, these effects have also been found in cortical areas (Wallace and others 1992; Laurienti and others 2002). The first main type of such effects is known as multisensory enhancement and is reflected in a significant enhancement for responding to multisensory stimuli, relative to the combined response activity to either unisensory stimulus alone. The second pattern, known as multisensory depression, is initiated when one sensory stimulus located outside its modality-specific receptive field degrades or eliminates the neuron's responses to another sensory stimulus presented within its modality-specific receptive field. Although somewhat lesser understood than multisensory enhancement, the latter pattern is also considered to be a key index of multisensory integration because it would tend to decrease the likelihood of multimodal stimuli presented at different locations from being integrated into one multisensory percept. Although this question has not been addressed yet,

it would appear logical that selective attention would be one factor that might be able to modulate this multisensory depression effect. Interestingly, the differences in selectively attending to either one or both modalities suggest that the depression effect may consist of separate mechanisms. Specifically, the depression effect in the unattended condition bears resemblance to the P50 amplitude effect that is commonly observed in sensory gating studies and that is believed to be a reflection of a more or less automatic mechanism involved in filtering out irrelevant inputs (Boutros and Belger 1999).

### ***The Relative Roles of Visual and Auditory Attention in Multisensory Integration***

The second main finding is that the early electrophysiological effects of multisensory integration appear to be critically dependent on the subject attending to both the visual and auditory modalities simultaneously. In the present study, this conclusion is based on 2 independent observations. First, it is only when both auditory and visual modalities are attended that subjects respond faster and more accurately to multisensory targets than to either of the unisensory targets alone. When only 1 of the 2 modalities was attended, RT and accuracy effects were either absent (auditory) or even negative (visual). As we have discussed previously (Talsma and Woldorff 2005a), the absence of such a pattern of behavioral improvement on multisensory stimuli could be indicative of subjects attempting to filter out sensory information from the irrelevant modality, instead of integrating this information with the stimuli present in the attended modality, which could then result in a behavioral cost in processing the multisensory stimuli. The latter pattern of results is, indeed, what was found when subjects were attending the visual modality. Second, the early superadditive effects on the AV - (A + V) difference waves were found to be significant only when subjects were attending to both the visual and the auditory modalities simultaneously.

To our knowledge, no study has yet investigated the specific contributions of visual or auditory attention alone on multisensory integration processes. Therefore, possible explanations for the observed effects necessarily remain somewhat speculative. It would seem plausible, however, that the integration of visual and auditory information is a process that is conducted most seamlessly when both visual and auditory modalities are attended. Many previous studies have investigated the roles of congruent versus incongruent auditory and visual information, such as irrelevant speech sounds on the interpretation of visual lip-reading patterns (McGurk and MacDonald 1976), the relative location of visual and auditory stimuli (Bertelson and others 2000; Vroomen and others 2001a, 2001b; Lewald and Guski 2003; Busse and others 2005; Teder-Sälejärvi and others 2005), and the effect of irrelevant sounds on the temporal order of a visual stimulus sequence (Shams and others 2001, 2002). These studies suggest that auditory and visual stimuli tend to be more or less integrated, even when parts of these stimuli are not task relevant. However, none of the studies mentioned above have reported either physiological or behavioral responses of such auditory or visual objects in the context of unimodal versus multimodal stimulus presentation using a manipulation of attending to one versus the other versus both modalities. Fort and others (2002a) investigated the effects of nonredundant target properties on multisensory integration processes and found some results that are consistent with the ones reported here: They reported not finding any early effects of multisensory

integration when target properties of the visual and auditory stimulus components were nonredundant. This result is somewhat similar to the present study, where the target properties of the unattended modality were also not beneficial for the task. Consequently, Fort and others found that responses to multisensory stimuli were neither faster nor more accurate when subjects were required to identify independent target properties in both visual and auditory modalities separately. This result contrasts with other studies from the same group (Giard and Peronnet 1999; Fort and others 2002b), which demonstrated that the simultaneous presentation of (attended) auditory and visual stimuli with redundant target features led to early electrophysiological effects as well as a behavioral improvement in detecting multisensory targets. Although the experimental manipulations used in the Fort and others (2002a) study are somewhat different than what was done in the present study, their results, just as ours, suggest that the multisensory integration effect only occurs early in time when both visual and auditory stimulus features are relevant and can be consistently constructed into a single coherent audiovisual object.

The results above could possibly be explained by a 2-stage multisensory integration mechanism, consisting of an inhibitory as well as a facilitatory component. In everyday life, our perceptual system is bombarded by a plethora of audiovisual stimuli, and it would seem logical that the default action of the brain is to inhibit responding to these stimuli, unless attention is directed to at least one stimulus feature, at which point the integrative processes for this stimulus are no longer inhibited but also not yet facilitated. According to this view, the full attention of both the visual and auditory systems (i.e., attending to both modalities) would be required to see a full facilitation of the early audiovisual integration processes.

It appears, however, that this mechanism is specifically involved in the earliest phases of multisensory integration. We found that amplitude of the anterior N1 was mainly larger for the multisensory stimuli than for the sum of the unisensory ones when both modalities were attended. Conversely, when the RSVP stream was attended (and the visual and auditory objects were unattended), no evidence for N1 differences between AV and [A + V] could be found, suggesting that the N1 generators are not affected by the initial inhibitory process but are affected by the multisensory integration effect when they are relevant. More research would be needed, however, to fully address this question.

### ***Supramodal Attention Effects on Audiovisual Stimuli***

Lastly, the present data indicate that visual attention can evoke a late auditory processing negativity in multisensory, but not in unisensory (visual or auditory), objects. Interestingly enough, no early effects of multisensory integration were found when subjects were attending the visual modality only. In addition, they were also slower and less accurate in processing multisensory stimuli than in processing unisensory visual stimuli in this condition. Therefore, one possibility would be that these late negativities are a reflection of a prolonging of a generic stimulus-processing mechanism, which occurs for multisensory but not for unisensory stimuli. However, this being the case, we would also have expected to find increases in the late negative waves, regardless of which modality was attended. If these late negativities were generated by a generic prolongation of processing mechanism in multisensory stimuli, these slow potentials elicited by the multisensory stimuli would have been

superimposed on the late processing negativities when the auditory modality was attended. For unisensory stimuli, these slow potentials would not be present and therefore would have shown up in the AV - [A + V] difference waves in all conditions.

Alternatively, these results indicate that although no active integration of auditory and visual stimuli occurs in the attend visual condition, at least at early stages of processing, some form of temporal binding did nevertheless occur at later stages, which triggers in a more specific way the evaluation of the auditory stimulus component of the multisensory stimulus. This conclusion would be consistent with our observation that we could find no evidence suggesting that the multisensory processing negativity would be generated by brain areas other than those generating the auditory processing negativity. Our finding of a late processing negativity-like component has similarities to a recent finding by Busse and others (2005). Using both ERP and functional imaging methods, they found differential processing of unattended auditory stimuli that was dependent on whether or not this stimulus was linked to a temporally co-occurring attended versus unattended visual object. It should be noted that there are several ways in which the effects reported by Busse and others (2005) are functionally different from the effects reported here. First, the auditory stimuli in the study of Busse and others (2005) were deliberately presented at different spatial locations from the temporally co-occurring visual ones (cf. ventriloquism effect). In addition, the temporal pairing of the auditory stimulus elicited an enhanced late negative-wave response when it was paired with an attended visual stimulus versus an unattended visual stimulus, whereas in the present study, this response was elicited by visually attended multisensory stimuli as compared with when the auditory stimulus was presented in isolation. Scalp topography plots of the ERP data of this "spreading of attention"-induced effect was similar to the topographies presented here. In addition, the functional magnetic resonance imaging data presented by Busse and others (2005) showed enhanced activations in auditory cortex to stimuli that were paired with an attended visual object relative to when paired with an unattended object. Thus, attending to the visual modality does apparently lead to an association of visual and auditory stimulus features and a spread of enhanced processing to the auditory components of the stimulus, even if this is not reflected in initial ERP activity or an immediate behavioral advantage.

### **Attention Effects**

In the present study, the focus of attention was manipulated by means of instructions at the start of each block of trials. These instructions were carried out successfully, as was evidenced by 1) the presence of typical attention effects in the ERPs, such as occipital P1 amplitude enhancements (Hillyard and others 1998), and 2) occipital selection negativities (Smid and others 1999) to attended visual stimuli, as well as frontocentral N1 enhancements following attended auditory stimuli (Näätänen and Picton 1987). These observations are in line with what is typically found in the literature and are therefore largely confirmatory that subjects did indeed successfully focus their attention. Of particular importance here, however, is the observation that attention effects elicited by multisensory stimuli consisted of a combination of effects typically elicited by visual and auditory stimuli (i.e., consisting of a combination of the occipital and frontocentral effects described above). This

observation would be consistent with prior results from the relatively limited number of studies that have examined the effects of selective attention on multisensory stimuli (Czigler and Balazs 2001; Talsma and Woldorff 2005a).

### **Summary and Conclusions**

The present study investigated the time course of ERP reflections of multisensory integration and their interactions with attention. Based on these results, the answers to the questions posed in the beginning of the present article are 1) that attending to stimuli at central locations does indeed lead to early effects of multisensory integration and 2) that it does indeed matter for integration effects whether attention is directed at the visual modality, the auditory modality, or both. More specifically, the data show that it is required to attend to both modalities to fully facilitate audiovisual integration. Three main findings of this study support these conclusions: 1) the early superadditive effect of integration for attended multisensory objects actually inverted to a subadditive effect when subjects were attending away from the objects; 2) early effects of multisensory integration, such as those reported previously in the ERP literature (Giard and Peronnet 1999; Fort and others 2002b; Molholm and others 2002), appeared to occur only when subjects were attending to both visual and auditory modalities simultaneously; and 3) despite the absence of early multisensory attention effects and despite a behavioral disadvantage in processing multisensory stimuli when subjects were attending the visual modality only, the auditory and visual stimuli eventually became associated with each other, as evidenced by the observation that visual attention can induce a late processing-like component on a multisensory stimulus but not on a visual stimulus alone or on an auditory stimulus alone.

We thus conclude that when attention is directed to both modalities simultaneously, auditory and visual stimuli are integrated very early in the sensory flow of processing (~50 ms poststimulus). Attention appears to play a crucial role in initiating such an early integration of auditory and visual stimuli. When only one modality is attended, the integration processes appear to be delayed. Nevertheless, even when only one modality is attended, the temporal co-occurrence of the stimulation in the 2 modalities will cause them to be associated with each other but at later stages of processing. Evidence for the latter conclusion was provided by the observation of a late processing negativity elicited by a multisensory stimulus, apparently reflecting late enhanced processing of the auditory component of the stimulus, even when attention was directed only to the visual component of this stimulus.

### **Notes**

This study was supported by National Institutes of Health grants R01-MH64015, National Science Foundation grant 0524031, and R01-NS051048 to MGW. We wish to thank Laura Busse, Tineke Grent't Jong, and Roy Strowd for their assistance during various stages of this project. We would also like to acknowledge the helpful comments from 2 anonymous reviewers on an earlier draft of this work. *Conflict of Interest:* None declared.

Address correspondence to Durk Talsma, Department of Cognitive Psychology, Vrije Universiteit, Van den Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Email: d.talsma@psy.vu.nl

### **References**

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.

- Boutros NN, Belger A. 1999. Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. *Biol Psychiatry* 45:917-922.
- Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG. 2005. The spread of attention across modalities and space in a multisensory object. *Proc Natl Acad Sci USA* 102(51):18751-18756.
- Calvert GA, Campbell R, Brammer MJ. 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649-657.
- Czigler I, Balazs L. 2001. Event-related potentials and audiovisual stimuli: multimodal interactions. *Neuroreport* 12:223-226.
- Driver J. 1996. Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381:66-68.
- Eimer M, Schröger E. 1998. ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology* 35:313-327.
- Fort A, Delpuech C, Pernier J, Giard MH. 2002a. Early auditory-visual interactions in human cortex during nonredundant target identification. *Cogn Brain Res* 14:20-30.
- Fort A, Delpuech C, Pernier J, Giard MH. 2002b. Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex* 12:1031-1039.
- Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE. 2000. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cogn Brain Res* 10:77-83.
- Giard MH, Peronnet F. 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473-490.
- Handy TC. 2005. Basic principles of ERP quantification. In: Handy TC, Editor. *Event-related potentials: a methods handbook*. 1st ed. Cambridge, MA: MIT Press. p 33-55.
- Hillyard SA, Mangun GR, Woldorff MG, Luck SJ. 1995. Neural systems mediating selective attention. *Handbook of cognitive neuroscience*. Cambridge, MA: MIT Press. p 665-681.
- Hillyard SA, Vogel EK, Luck SJ. 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos Trans R Soc Lond Ser B Biol Sci* 353:1257-1270.
- Hopfinger JB, Mangun GR. 1998. Reflexive attention modulates processing of visual stimuli in human extrastriate cortex. *Psychol Sci* 9: 441-447.
- Jiang W, Stein BE. 2003. Cortex controls multisensory depression in superior colliculus. *J Neurophysiol* 90:2123-2135.
- Kenemans JL, Kok A, Smulders FTY. 1993. Event-related potentials to conjunctions of spatial-frequency and orientation as a function of stimulus parameters and response requirements. *Electroencephalogr Clin Neurophysiol* 88:51-63.
- Kenemans JL, Lijffijt M, Camfferman G, Verbaten MN. 2002. Split-second sequential selective activation in human secondary visual cortex. *J Cogn Neurosci* 14:48-61.
- Laurienti PJ, Burdette JH, Wallace MT, Yen Y, Field AS, Stein BE. 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. *J Cogn Neurosci* 14:420-429.
- Lewald J, Guski R. 2003. Cross-modal perceptual integration of spatially and temporarily disparate auditory and visual stimuli. *Cogn Brain Res* 16:468-478.
- Macaluso E, Eimer M, Frith CD, Driver J. 2003. Preparatory states in crossmodal spatial attention: spatial specificity and possible control mechanisms. *Exp Brain Res* 149:62-74.
- 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.
- McGurk H, MacDonald J. 1976. Hearing Lips and Seeing Voices. *Nature* 264:746-748.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cogn Brain Res* 14:115-128.
- Motter BC. 1993. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol* 70:909-919.
- Näätänen R, Picton T. 1987. The N1 wave of the human electric and magnetic response to sound: a review and analysis of the component structure. *Psychophysiology* 24:375-424.
- Schröger E, Giard MH, Wolff C. 2000. Auditory distraction: event-related potential and behavioral indices. *Clinical Neurophysiol* 111:1450-1460.
- Senkowski D, Talsma D, Herrmann CS, Woldorff MG. 2005. Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp Brain Res* 166:411-426.
- Shams L, Kamitani Y, Shimojo S. 2002. Visual illusion induced by sound. *Cogn Brain Res* 14:147-152.
- Shams L, Kamitani Y, Thompson S, Shimojo S. 2001. Sound alters visual evoked potentials in humans. *Neuroreport* 12:3849-3852.
- Smid HGOM, Jakob A, Heinze HJ. 1999. An event-related brain potential study of visual selective attention to conjunctions of color and shape. *Psychophysiology* 36:264-279.
- Stein BE, London N, Wilkinson LK, Price DD. 1996. Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J Cogn Neurosci* 8:497-506.
- Stein BE, Meredith MA. 1993. *The merging of the senses*. Cambridge, MA: MIT press.
- Stein BE, Wallace MT. 1996. Comparisons of cross-modality integration in midbrain and cortex. *Prog Brain Res* 112:289-299.
- Talsma D, Kok A. 2001. Nonspatial intermodal selective attention is mediated by sensory brain areas: evidence from event-related potentials. *Psychophysiology* 38:736-751.
- Talsma D, Kok A. 2002. Intermodal spatial attention differs between vision and audition: an event-related potential analysis. *Psychophysiology* 39:689-706.
- Talsma D, Woldorff MG. 2005a. Attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J Cogn Neurosci* 17:1098-1114.
- Talsma D, Woldorff MG. 2005b. Methods for the estimation and removal of artifacts and overlap in ERP waveforms. In: Handy TC, Editor. *Event-related potentials: a methods handbook*. Cambridge, MA: MIT Press. p 115-148.
- Teder-Sälejärvi W, Di Russo F, McDonald JJ, Hillyard SA. 2005. Effects of spatial congruity on audio-visual multimodal integration. *J Cogn Neurosci* 17:1396-1409.
- Tootell RBH, Hadjikhani NK, Mendola JD, Marrett S, Dale AM. 1998. From retinotopy to recognition: fMRI in human visual cortex. *Trends Cogn Sci* 2:174-183.
- Van der Heijden AHC. 1992. *Selective attention in vision*. London: Routledge.
- Van der Heijden AHC. 1993. The role of position in object selection in vision. *Psychol Res* 56:44-58.
- Vroomen J, Bertelson P, De Gelder B. 2001a. Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychol* 108:21-33.
- Vroomen J, Bertelson P, De Gelder B. 2001b. The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept Psychophys* 63:651-659.
- Wallace MT, Meredith MA, Stein BE. 1992. Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91:484-488.
- Wallace MT, Stein BE. 1997. Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci* 17:2429-2444.
- Wallace MT, Stein BE. 2001. Sensory and multisensory responses in the newborn monkey superior colliculus. *J Neurosci* 21:8886-8894.
- Woldorff MG. 1993. Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction. *Psychophysiology* 30:98-119.
- Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE. 1993. Modulation of early sensory processing in human auditory-cortex during auditory selective attention. *Proc Natl Acad Sci USA* 90:8722-8726.
- Woldorff MG, Hackley SA, Hillyard SA. 1991. The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology* 28:30-42.