Psychophysiology, 43 (2006), 541–549. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2006.00452.x

Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality

DURK TALSMA,[a,b] TRACY J. DOTY,[a] ROY STROWD,[a] AND MARTY G. WOLDORFF[a]

[a]Center for Cognitive Neurosciences, Duke University, LSRC-Bldg, Box 90999, Durham, North Carolina, USA
[b]Cognitive Psychology Department, Vrije Universiteit, Van den Boechorstraat, 1, 1081 BT Amsterdam, the Netherlands

Abstract

One finding in attention research is that visual and auditory attention mechanisms are linked together. Such a link would predict a central, amodal capacity limit in processing visual and auditory stimuli. Here we show that this is not the case. Letter streams were accompanied by asynchronously presented streams of auditory, visual, and audiovisual objects. Either the letter streams or the visual, auditory, or audiovisual parts of the object streams were attended. Attending to various aspects of the objects resulted in modulations of the letter-stream-elicited steady-state evoked potentials (SSVEPs). SSVEPs were larger when auditory objects were attended than when either visual objects alone or when auditory and visual object stimuli were attended together. SSVEP amplitudes were the same in the latter conditions, indicating that attentional capacity between modalities is larger than attentional capacity within one and the same modality.

Descriptors: EEG, Electrophysiology, Intermodal attention, Oscillations, Multisensory processing

The ability of the human mind to focus its attention on specific aspects of the outside world is believed to be brought about in the brain by cortical and subcortical circuits that are able to send modulatory signals to perceptual areas, which then selectively modify the sensitivity of neurons involved in the perceptual processing of these aspects (Kastner & Ungerleider, 2000; LaBerge, 1995). There is evidence that this modulation consists of an increase in sensitivity of neurons that are responsive to the attended feature, in combination with a simultaneous decrease in sensitivity of neurons responsive to nonattended features (Hopfinger, Buonocore, & Mangun, 2000; Motter, 1993; Woldorff, Hackley, & Hillyard, 1993).

The observation that selection of relevant stimuli takes place at the perceptual stage has been interpreted as evidence for the notion that these perceptual processes are limited in capacity (Kastner & Ungerleider, 2000). Although this notion of a perceptual limitation is currently largely agreed upon for concurrent streams of stimuli that are presented within the same sensory modality, it is still largely debated whether the same degree of limitation exists when the brain has to focus attention on one of two or more concurring stimuli presented in different modalities, such as in vision and audition (Arnell & Jolicœur, 1996; Duncan, Martens, & Ward, 1997; Eimer & Schröger, 1998; Escera, Alho, Schröger, & Winkler, 2000; Jolicœur, 1999; Schröger, Giard, & Wolff, 2000; Talsma & Kok, 2001, 2002).

It is generally believed that the control areas from which the biasing signal originate can be found in frontal and parietal brain areas, specifically regions in a dorsal fronto-parietal network (Banich et al., 2000; Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Hopfinger et al., 2000; LaBerge, 1995; Woldorff et al., 2004; for a review, see Corbetta & Shulman, 2002). Interestingly, some recent fMRI studies have found some overlap in brain area activation to be involved in spatial and nonspatial visual attentional orienting (Giesbrecht, Woldorff, Song, & Mangun, 2003). In addition, a recent fMRI study has suggested that the parietal part of this network is also involved in the control of auditory attention (Shomstein & Yantis, 2006). Lastly, several event-related potential (ERP) studies have shown that attending to visual locations in space also enhances physiological brain responses to auditory stimuli at that location and vice versa (e.g., Eimer & Schröger, 1998; Hillyard, Simpson, Woods, Van Voorhis, & Münte, 1984; Talsma & Kok, 2002).

These results have therefore suggested that the attentional control areas are supramodal in nature; that is, there is a single network of brain areas involved in the allocation of attentional capacity. It is still a matter of ongoing debate, however, whether or not this is fully the case. It is still not clear whether there is one

---

This study was supported by NIH grants R01-MH64015, NSF grant 0524031, and R01-NS051048 to M.G.W. We thank Laura Busse and Tineke Grent-'t Jong for their assistance during data collection and Hein van Schie and Sander Los for their helpful comments on an earlier version of the manuscript. In addition, we also thank Dr. John J. McDonald and one anonymous reviewer for helpful comments on an earlier draft of this article.

T.J.D. is now at the Laboratory of Brain and Cognition, NIMH, Bethesda, MD, USA and the Department of Clinical Neuroscience, Karolinska Institutet, Karolinska Hospital, Stockholm, Sweden.

Address reprint requests to: Durk Talsma, Cognitive Psychology Department, Vrije Universiteit, Van den Boechorststraat 1, 1081BT Amsterdam, The Netherlands. E-mail: d.talsma@psy.vu.nl.

single control system modulating the neural sensitivity of the visual and auditory perceptual areas or whether this biasing is done by independent control resources for the visual and auditory modalities. For example, LaBerge (1995) has made a distinction between attentional orienting and attentional resolving. Whereas attentional orienting is mainly involved in shifting attention between locations and/or features, the attentional resolving mechanism is believed to extract relevant stimulus features from the attended object stream. It is still largely unclear whether the latter attention-related resolving mechanisms are supramodal in nature or not. The present study seeks to answer this question.

The effects of the aforementioned sensitivity changes of the sensory brain areas are generally expressed in differences in the amplitude of neural responses to task-relevant stimuli. These can be recorded as amplitude changes in ERP components, as well as by modulations of the responsiveness of a steady-state visual evoked potential (SSVEP). A SSVEP can be recorded as a continuous, driven, oscillatory response over the visual cortex by repeatedly presenting a visual stimulus at a constant rate, which, accordingly, gives it the same fundamental frequency as the triggering stimulus sequence. Moreover, it has previously been demonstrated that the amplitudes of these steady-state waves are substantially enlarged when the evoking stream is attended (Müller & Hillyard, 2000; Müller, Picton, et al., 1998; Müller, Teder-Sälejärvi, & Hillyard, 1998), compared to when it is unattended.

In the present study, we applied this approach by combining a rapid serial visual presentation (RSVP) paradigm with a multiple-stream selective attention task (see also Talsma, Doty, & Woldorff, in press). This was done to determine whether or not attending to concurrent events that were presented either within or outside of the visual modality would differentially affect the amplitude of the SSVEP elicited by the RSVP stream. To accomplish this, we presented a rapidly changing stream of letters while simultaneously presenting additional asynchronously presented objects in the visual and auditory modalities. During separate runs of trials, participants were instructed to attend to either the letter stream or the objects. In the latter conditions, they were instructed either to attend to the visual or auditory objects separately, or attend to both auditory and visual objects simultaneously.

Based on previous findings (Müller & Hillyard, 2000; Müller, Picton, et al., 1998; Müller, Teder-Sälejärvi, et al., 1998), we expected that the letter-stream SSVEP amplitude would be larger when that stream was attended, in comparison to the other conditions. More importantly for the present goals, for the other three conditions we predicted different outcomes depending on whether or not the auditory and visual attentional resolving mechanism has a central, amodal bottleneck limitation. If it does, we would expect that the letter-stream SSVEP amplitude would be the lowest when attention was directed to *both* the visual and auditory object streams, because in this condition, the least attentional resolving capacity would be left for processing the RSVP items. Furthermore, we would then predict that in comparison to this condition, the SSVEP would be of a somewhat higher amplitude for the conditions in which *only* the visual or auditory object streams were attended, but that the SSVEP responses would be *equal* in these two conditions.

If, on the other hand, attentional capacity is more restricted within a single modality than across the visual and auditory modalities, we would expect that the SSVEP amplitude elicited by the RSVP letter stream would be larger when the concurrently presented auditory objects were attended than when the concurrently presented visual objects were attended. Furthermore, we would predict to find no SSVEP amplitude differences between the attend-visual and attend-audiovisual conditions, because the added demands of attending to auditory objects would not further decrease the responsiveness of the visual brain areas.

Methods

Participants

Written informed consent was obtained from each of the 19 volunteers who participated in the experiment (age 18–25, mean age 19). Of these, 8 were female and 2 were left-handed. All participants had normal or corrected-to-normal visual capabilities and normal hearing. None of them reported having a history of mental or physical illness.

Stimuli and Procedure

A random stream of letters (1 × 1°) was presented at an eccentricity of 1.5° above fixation on a computer screen at a rate of 6.67 Hz (150 ms per letter). Randomization was restricted to prevent two identical letters from being presented in succession, which would have caused a distracting break in apparent rate. At random intervals, between 1 and 10 s, a digit was embedded in the stream, instead of a letter, serving as a target stimulus in one of the experimental conditions. Concurrent with this stream, at random stimulus onset asynchrony (SOA) intervals of 350 to 650 ms, additional unisensory visual, unisensory auditory, and multisensory audiovisual objects were presented. The visual stimuli consisted of square-wave horizontal line gratings (subtending an angle of 5° and consisting of four white bars against a black background) that were presented with a 105-ms duration, at an angle of about 3° directly below fixation (see Figure 1). Auditory stimuli consisted of 1600-Hz tone pips (65 dB(A) SPL), also with a duration of 105 ms and with linear rise and fall time times of 10 ms each. These tones were presented from speakers placed directly behind the monitor, such that the subjective location of the auditory stimuli matched that of the visual objects (Eimer & Schröger, 1998). Audiovisual objects consisted of the simultaneous presentation of the visual and auditory stimulus components.

Preceding the start of each block of trials, participants were instructed to direct their attention to one of four possible stimulus-types combinations: (1) *Attend RSVP*: subjects were instructed to focus their attention on the RSVP letter stream and to detect and respond to the target digits. (2) *Attend audio-visual*: Subjects were instructed to attend to both the visual and auditory object stimuli and to detect occasional targets in both the visual and auditory stimulus streams. Target stimuli were highly similar to standards, but contained a transient dip in intensity half-way 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, Talsma, Herrmann, & Woldorff, 2005; Talsma & 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 the visual components of the multisensory objects) to detect visual

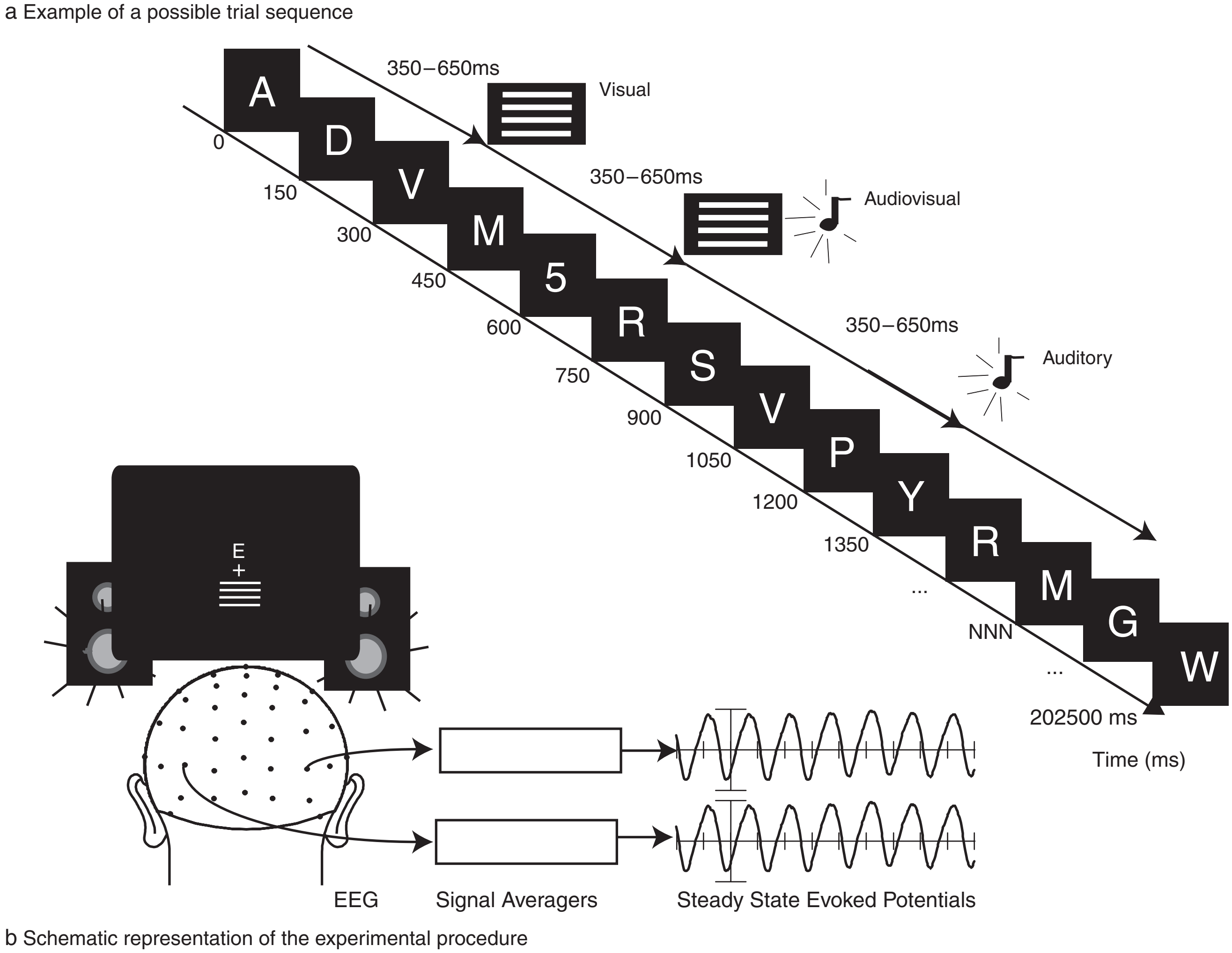

Figure 1. Illustration of the paradigm. A: Example of a possible sequence of letters and objects used in the experiment. The update frequency of the letters was fixed at 6.7 Hz (150 ms per letter), and the letter sequence was randomized, except for the constraint that any one letter could never be presented twice in a row. Stimulus onset asynchronies (SOAs) of the concurrent auditory and visual object stimuli were randomized between 350 and 650 ms. All three object stimulus types (visual, auditory, and multisensory) were presented in a randomized, first-order counterbalanced sequence in each block of trials. B: Schematic representation of the placement of the letter stream and the visual, and auditory stimulus objects. While a regular stream of letters was presented above a central fixation dot, an irregular stream of visual and auditory stimuli was presented below fixation. Speaker placement was immediately behind and somewhat below the center of the monitor, so that the auditory stimuli were subjectively perceived as originating from the same location as the visual objects.

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) to detect auditory targets among these.

In each of the above-described conditions, there were always 20% visual targets, 20% auditory targets, and 20% multisensory targets. However, in the attend-visual condition, only the visual and multisensory targets were behaviorally relevant, and in the attend-auditory condition, only the auditory and multisensory targets were relevant. Participants were required to respond to these targets by making a speeded manual response.

To ensure an adequate difficulty of these targets and to familiarize participants with the stimulus material, each participant completed a unisensory target discrimination task in several practice runs at the start of the experimental session, in which the difficulty of the visual and auditory targets was independently adjusted to each participant's individual ability (Talsma & Woldorff, 2005a), such that each participant would be able to discriminate 90% of all the targets. In the auditory practice run, participants were presented randomly (50% probability) either a standard tone or a target (i.e., a stimulus with a midduration intensity decrement) and they were required to indicate whether the stimulus was a standard or a target. Based on the subject's accuracy, the difficulty of the presented target was changed by increasing or decreasing the level of the midstimulus intensity decrement. More specifically, if a subject's accuracy was below 90% correct, target difficulty was decreased by increasing the midduration intensity decrement (i.e., making the decrement larger and more discriminable), and when subject's accuracy was above 90% correct, target difficulty was increased by decreasing the midduration intensity decrement, thus making standards and targets more similar. For the visual stimuli, a similar procedure was used. Based on previous studies (Senkowski et al., 2005;

Talsma & Woldorff, 2005a) we expected that accuracy would fall to about 80% correct in the somewhat more demanding audiovisual conditions. These procedures ensured that the task was difficult but doable, helping in turn to ensure that the participant's attention was fully focused on that task. To determine whether or not visual and auditory object target stimuli were equally difficult to detect, mean reaction times and hit rates were computed for the unisensory visual and auditory objects in the attend-auditory, attend-visual (single modality), and attend-audiovisual (attending both modalities simultaneously) conditions. Responses were considered correct if a response followed the relevant target within in a time window of 100 to 1500 ms. Reaction times and hit rates were analyzed separately, using pairwise t tests. Reaction times extending beyond two standard deviations from the mean were excluded from the analyses. In addition, reaction times and error rates to targets in the RSVP stream are reported.

Each block of trials took about 4 min to complete and consisted of 100 visual, 100 auditory, and 100 audiovisual objects. The letter stream was concurrently presented during the entire length of the trial block, resulting in the presentation of approximately 1350 letters per block. The exact number of letters presented during each block varied slightly, due to the random and variable SOA between the object stimuli. For each attention condition, participants completed two blocks of trials. The order of the attention conditions was counterbalanced across participants. Furthermore, each participant's response hand was counterbalanced across blocks of trials, such that they completed one block of each condition with their left hand and one with their right hand, for each attention condition.

EEG Recordings

Stimulus presentation was controlled by a personal computer running the "Presentation" software package (Neurobehavioral Systems, Inc., Albany, CA). ERP recordings were recorded from 64 equally spaced tin electrodes (FPz, Fz, FCz, Cz, Fp1m, Fp2m, F3a, F4a, F3s, F4s, FC1, FC2, C1a, C2a, F7a, F8a, F3i, F4i, C3a, C4a, PA1a, PA2a, F7p, F8p, C5a, C6a, T3′, T4′, LC, RC, LIO, RIO, Lm, Inz, Ozi, Ozs, Pzi, Pzs, I1, I2, O1i, O2i, O1′, O2′, PO1, PO2, P1′, P2′, C1p, C2p, TI1, TI2, TO1, TO2, P3i, P4i, P3a, P4a, C3′, C4′, T35i, T46i, C5p, and C6p)[1] mounted in a custom-designed elastic cap (Electro-Cap International, Inc., Eaton, OH) and referenced to the right mastoid during recording. 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 two electrodes at the outer canthi of the eyes. Vertical eye movements and eyeblinks were detected by electrodes placed below the orbital ridge of both eyes, which were referenced to two electrodes located directly above the eyes. During recording, eye movements were also monitored using a closed circuit video monitoring system. EEG was recorded using a Neuroscan (SynAmps) acquisition system using a bandpass filter of 0.01 to 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: SSVEPs

For each of these four conditions, the SSVEPs evoked by the RSVP letter streams were calculated in a manner using a time-invariant frequency analysis similar to that described by Müller, Picton, et al. (1998). In each condition, a 2000-ms window (containing a prestimulus baseline period of 950 ms and a poststimulus period of 1050 ms), time-locked to the onset of each letter, was moved in steps of 150 ms, and these epochs were averaged across the entire length of each 4-min run, resulting in approximately 1350 steps for each average. A relatively long prestimulus baseline period was included in this step to optimize artifact detection procedures (see Talsma & Woldorff, 2005b). Epochs that were contaminated by ocular or bodily movement artifacts were excluded from the averaging procedure. After averaging, all channels were re-referenced to the algebraic average of the two mastoid electrodes.

The 1050-ms poststimulus part of the resulting waveforms represented seven cycles of the SSVEP. Figure 2 shows the poststimulus part of these waveforms from one representative participant. Although these waveforms provide a good estimate of each individual participant's SSVEPs, phase variations between participants can be considerable, causing cross-participant averages of these waveforms to underestimate the attention effects on these SSVEP waveforms (Müller, Picton, et al., 1998). To obtain amplitude values of the SSVEP, these signals were subjected to a fast Fourier transform (FFT), which was performed for each participant separately. To minimize spectral leakage, and because the FFT computation requires that the input number of data points be equal to a power of two, the 1050-ms poststimulus time window (representing seven complete cycles of the SSVEP) was up-sampled, using a spline interpolation algorithm, to 2048 data points before being subjected to the FFT procedure (see Figure 2). This method ensured that the 6.7-Hz frequency was at the center of one of the FFT bins. Considering that the original data were already sampled at a fairly high digitization rate (500 Hz) relative to the low-pass analog filter setting used during recording (< 100 Hz), and that the data were further digitally low-pass filtered prior to the FFT (< 56 Hz), it seems rather unlikely that this resampling of the data could have led to any significant interpolation errors. The SSVEP spectral amplitude was extracted from FFTs by obtaining the modulus of the real and imaginary part at each frequency, at each electrode. The estimates for the 6.7-Hz wave obtained from the lateral occipital electrodes TO1 and TO2 were subjected to a within-subjects ANOVA. To estimate the scalp topography of the SSVEP activity, the SSVEP amplitude estimates at each site for the attend-audiovisual condition were subtracted from those obtained in the attend-RSVP stream, attend-auditory, and attend-visual conditions. These differences were then plotted as scalp topographies.

Results

Behavioral Data

Reaction times. For each condition where attention was directed toward the auditory and visual objects (i.e., attend-visual, attend-auditory, and attend-audiovisual) the reaction times to

---

[1] These electrode positions are named relative to their approximate 10-10 equivalents. A suffix of "s" indicates that the electrode was placed slightly (i.e., within 1–1.5 cm) superior to the indicated standard position; "i" indicates it was placed inferior to the standard position. Similarly, "a" and "p" indicate the electrode in question was positioned slightly anterior or posterior to the standard locations, respectively. Electrodes position within 0.5–1.0 cm are named by the standard location name with a prime mark (e.g., C4′).

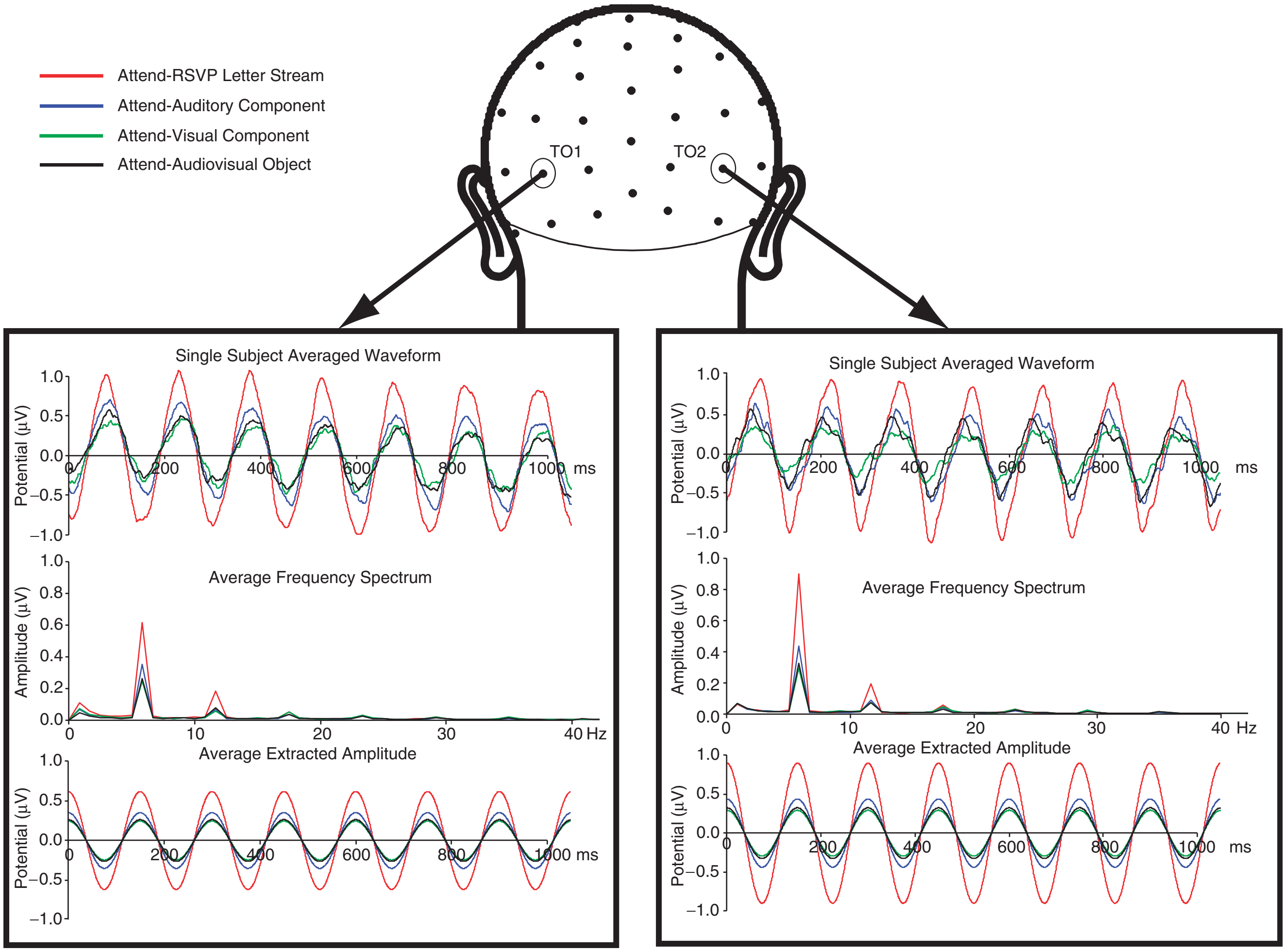

Figure 2. SSVEP effects at electrodes TO1 and TO2. Shown for each electrode, from top to bottom, are (1) the amplitude effects of the SSVEP obtained by selective averaging for one representative participant. Although the preaveraged data provide a good estimate of each single participant's SSVEP, grand averages across participants tend to underestimate these effects, due to phase variation in SSVEP amplitudes between participants (see text). (2) Average power spectrum obtained by submitting the preaveraged data for each participant (as in 1) to the FFT procedure and averaging the single subject power spectra. As can be seen, the SSVEPs elicited by the letter streams were smallest when participants were attending to a concurrently presented visual or audiovisual object and largest when they were attending the letters. Critically, the SSVEP amplitude was somewhat in between when participants were attending to a concurring auditory object only. (3) Grand-average of the phase-corrected grand-average 6.7-Hz waveform.

visual-only and to auditory-only objects were compared (1) to each other and (2) to the reaction times to audiovisual objects, using pairwise *t* tests (see Table 1). In addition, the reaction times to the unisensory stimuli were tested to check whether they differed between the focused (i.e., attend-visual or attend-auditory condition) and divided attention conditions (i.e., the attend-audiovisual condition).

In the two unisensory attention conditions (attend-auditory and attend-visual), the reaction times to the respective unisensory stimuli did not differ significantly from each other, $t(18) = 0.1$, $p > .9$. Although Table 1 suggests that responses to auditory targets were somewhat slower than responses to visual targets in the attend audiovisual condition, this effect failed to reach significance, $t(18) = 1.96$, $p < .06$.

In the attend-auditory condition, no significant reaction time difference between the auditory and audiovisual object targets could be found, $t(18) = 1.4$, $p > .1$. In the attend-visual condition, reaction times to visual-only object target were significantly faster than reaction times to audiovisual object targets, $t(18) = 4.18$, $p < .001$. In the attend-audiovisual condition, no significant difference could be found between visual and audiovisual target detection, $t(18) = 1.07$, $p > .3$, but reactions were significantly slower to auditory than to audiovisual targets, $t(18) = 4.6$, $p < .0005$. Mean reaction time to the target digits in the RSVP condition was 521 ms.

Responses to unisensory auditory stimuli were significantly faster in the attend-auditory condition than in the attend-audiovisual condition, $t(18) = 3.76$, $p < .005$. The difference between response times to unisensory visual stimuli in the attend-visual condition versus the attend-audiovisual condition was only marginally significant, however, $t(18) = 1.98$, $p < .06$.

Hit rates. Hit rates are given in Table 2. Response accuracy to unisensory visual and auditory stimuli differed in neither the unisensory attention conditions, $t(18) = 0.89$, $p > .38$, nor the attend-audiovisual condition, $t(18) = 0.24$, $p > .81$. In the attend-visual condition, responses to visual-only stimuli were more accurate than those to multisensory stimuli, $t(18) = 4.20$, $p < .0005$.

Table 1. *Mean Response Times in Milliseconds to the Target Stimulus Object Types*

	Stimulus		
	Auditory	Visual	Audiovisual
Att. Auditory	495 (74)		484 (59)
Att. Visual		493 (44)	541 (57)
Att. Audiovisual	544 (85)	511 (85)	496 (62)

Notes: Standard deviation is shown in parentheses. Empty cells indicate that participants were not supposed to respond to the respective combination of condition and stimulus type.

Similarly, in the attend-audiovisual condition, responses to audiovisual stimuli were slightly more accurate than those to either visual, $t(18) = 2.21$, $p < .05$, or auditory stimuli, $t(18) = 3.25$, $p < .005$, alone. Mean hit rate to the RSVP target digits in the attend-RSVP condition was 79%.

Responses to unisensory auditory target stimuli were significantly more accurate in the attend-auditory condition than in the attend-audiovisual condition, $t(18) = 6.6$, $p < .000005$. Similarly, the accuracy to unisensory visual target stimuli was significantly better in the attend-visual condition than in the attend-audiovisual condition, $t(18) = 3.16$, $p < .005$.

SSVEP Effects

The SSVEPs evoked by the letter stream under the four attention conditions are shown in Figure 2. This figure shows that the SSVEPs elicited by the letter stream when that stream was attended appeared to have the largest amplitude, as expected, followed by the SSVEPs during the attend-auditory condition, and then during the attend-visual and attend-audiovisual conditions. For statistical analysis, the SSVEP amplitudes evoked by the letter stream under the four attention conditions (see Figure 2) were determined by means of extracting the amplitude of the 6.7-Hz signal component, using a FFT of activity over the temporo-occipital lobe. A frequency of 6.7 Hz was used because that frequency corresponds with the presentation frequency of the letter stream evoking the SSVEP. These amplitudes were then subjected to ANOVA, with the two within-subject factors of Condition (four levels: attend-RSVP, attend-auditory, attend-visual, and attend-audiovisual) and Hemisphere (electrodes: TO1 and TO2). Greenhouse–Geisser correction was applied for the tests including the factor Condition. An overall main effect of Condition, $F(3,54) = 18.2$, $p < .0001$, indicated that the SSVEP amplitudes differed significantly across the four conditions (Figure 2).

We had expected that the SSVEP amplitude would be largest in the RSVP condition, smallest in the attend-audiovisual conditions, and somewhat intermediate in the attend-visual and attend-auditory conditions. Importantly, for the latter two conditions we had predicted two alternative patterns of results depending on whether or not attentional capacity was larger across modalities than within one and the same modality. Figure 2 indeed suggests that attentional capacity is larger across modalities. More specifically, whereas the SSVEP amplitude did not appear to differ between the attend-visual condition and the attend-audiovisual condition, it did appear to be larger in the attend-auditory condition than in either the attend-visual or attend-audiovisual condition. These amplitude differences were further investigated using three planned comparisons, contrasting the RSVP SSVEP amplitudes obtained in (1) the attend-RSVP versus attend-auditory conditions, (2) the attend-auditory versus attend-visual conditions, and (3) the attend-visual versus attend-audiovisual conditions.

The analyses confirmed that the letter-stream SSVEPs were larger in the attend-RSVP condition relative to the attend-auditory condition, $F(1,18) = 12.89$, $p < .005$, the condition with the next largest SSVEP amplitude. More interestingly, the SSVEP amplitude was also significantly larger in the attend-auditory condition than in the attend-visual condition, $F(1,18) = 7.00$, $p < .02$. Finally, the SSVEP amplitude did not differ between the attend-visual and attend-audiovisual conditions, $F(1,18) < 1$. Thus, this pattern of results supports the hypothesis that attentional capacity is larger across modalities than within a modality.

To illustrate that the attention effects on the SSVEP wave were the strongest over the visual brain areas, scalp topography plots were computed based on the 6.7-Hz FFT results from each scalp site. Figure 3B shows a bilateral increase in 6.7-Hz activity in the attend-auditory condition in comparison to the attend-audiovisual condition. A similar but stronger increase in 6.7-Hz activity can be seen in the attend-RSVP condition relative to the attend-audiovisual condition (or to any of the other RSVP-irrelevant conditions). Although Figure 3C suggests that this increase was somewhat stronger over the right hemisphere, no statistical evidence could be found for this laterality trend.

Discussion

The most novel finding in the present study is that the amplitude of a SSVEP evoked by a repeating letter stream was significantly larger when concurrent auditory objects were attended (attend auditory) than when other concurrent visual objects were attended (i.e., the attend-visual and attend-audiovisual conditions). Moreover, we could not find a difference in the SSVEP between the latter two conditions. This combination of findings thus rules out the possibility that there is a single, amodal, central capacity limit of attention.

The increase in SSVEP amplitude found in the attend-RSVP condition, relative to the other three conditions, is consistent with earlier studies employing SSVEPs (e.g., Müller, Picton, et al., 1998). One interesting difference between the present study and earlier work by Müller et al. is that we used a somewhat slower rate of presentation, resulting in a relatively low frequency SSVEP (6.7 Hz here vs. 20–28 Hz in the Müller et al. studies). These results therefore also demonstrate that the attentional enhancement of neural responses is effective over a relatively wide frequency range, showing a relative tonic increase in responsiveness in visual cortex that is modulated by attention. This tonic increase in responsiveness is also consistent with a relative phasic (Müller, Teder-Sälejärvi, et al., 1998) or tonic enhancement

Table 2. *Percentage Correctly Reported Target Stimulus Objects*

	Stimulus		
	Auditory	Visual	Audiovisual
Att. Auditory	84%		86%
Att. Visual		79%	61%
Att. Audiovisual	71%	70%	80%

Note: Empty cells indicate that participants were not supposed to respond to the respective combination of condition and stimulus type.

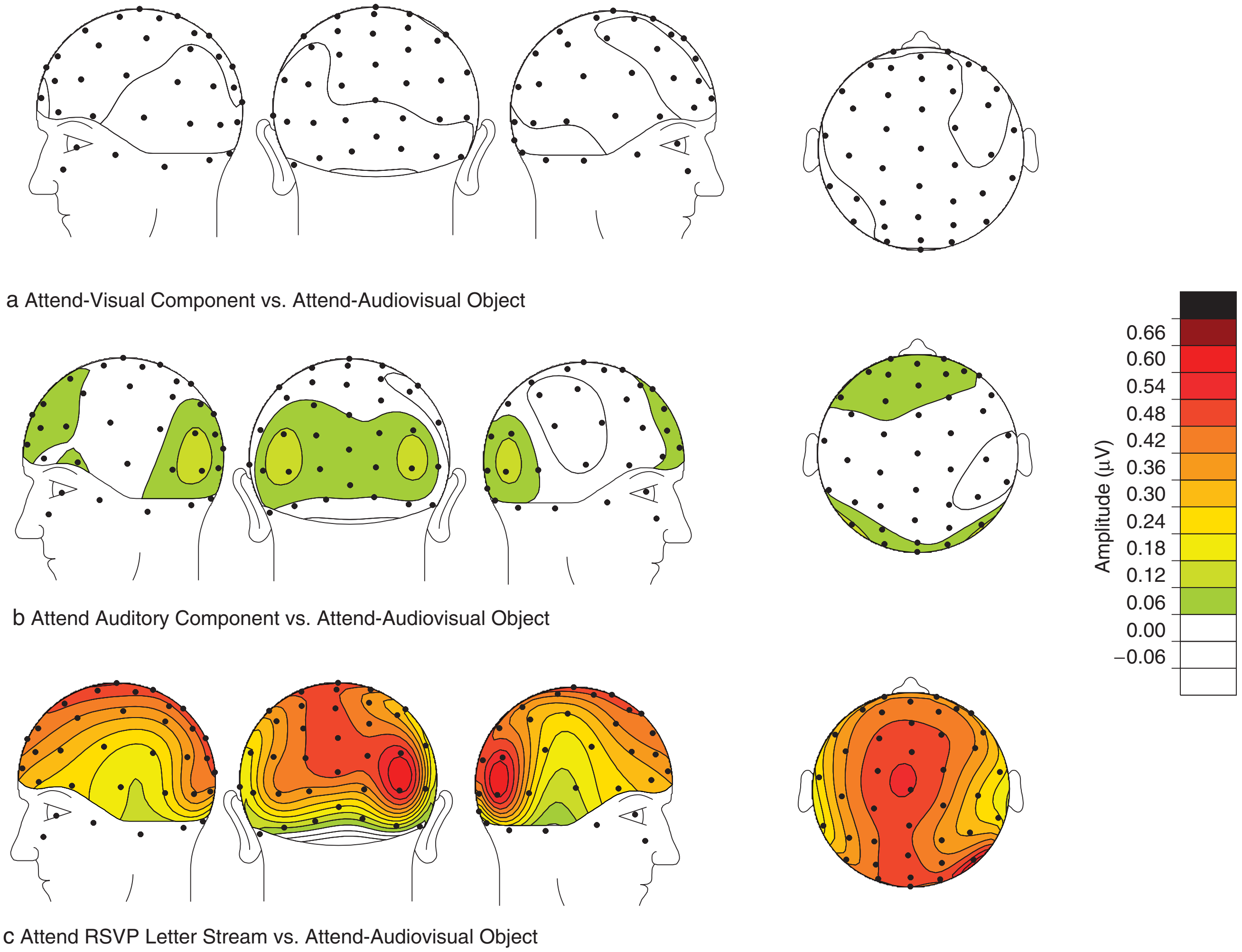

Figure 3. Spline-interpolated scalp topography maps of the amplitude differences between the four attention conditions. These topography maps are based on a grand average of the 6.7-Hz FFT frequency components. A: Comparison of the SSVEP amplitude differences between the attend-visual and attend-audiovisual objects conditions. No significant differences were found in this contrast. B: However, the temporo-occipital SSVEP waves were significantly enhanced in amplitude in the attend-auditory objects condition relative to the attend-audiovisual condition (or relative to the attend-visual condition, which elicited responses similar to the attend-audiovisual condition). C: Scalp topography of the large SSVEP waves in the attend-RSVP condition, also compared to the attend-audiovisual condition.

(Müller, Malinowski, Gruber, & Hillyard, 2003) that was established using a slightly different (i.e., time-varying) SSVEP analysis method.

Although the general pattern of results of the present study are in agreement with earlier studies reporting that the attentional blink phenomenon was more restricted within as compared to between modalities (Duncan et al., 1997), there is still an ongoing debate about whether attentional capacity is really more limited within than between modalities. Arguing against this conclusion is the observation that the processing of a simple auditory task could cause substantial interference in a concurrent visual encoding task (Arnell & Jolicœur, 1999; Jolicœur, 1999). The latter finding would be in agreement with the observation that irrelevant auditory stimuli can substantially distract attention from visual stimuli (Escera et al., 2000; Schröger et al., 2000). In contrast to these findings, however, it has been argued that attention is a hierarchical mechanism, which selects or rejects stimuli on the basis of the sensory system that is attended before conducting other types of selection (de Ruiter, Kok, & van der Schoot, 1998; Heslenfeld, Kenemans, Kok, & Molenaar, 1997; Talsma & Kok, 2001). Finally, yet another series of results has indicated that the processing of unattended visual stimuli can be affected by attending to the location of auditory stimuli and vice versa (Eimer & Schröger, 1998; Eimer & van Velzen, 2002; Talsma & Kok, 2002). The latter result would suggest that—should a common supramodal attentional system exist—it would operate on modulating the sensitivity of the visual and auditory systems in parallel.

The results from the present study imply that attentional capacity is more limited when relevant stimulus features have to be resolved among competing stimulus streams presented within one and the same modality as compared to when these competing stimulus streams are presented in different modalities. In other words, attending to an additional auditory stream does not affect the processing capacity of the visual stimuli as much as attending to another visual source would. These observations would therefore imply that attentional resolving (LaBerge, 1995) is a process that that has modality-specific capacity resources that are available, perhaps because it may be carried out largely by perceptual brain areas, relatively independently for visual and auditory stimulus streams. Taken together, these results suggest a distinction between the control systems involved in orienting and maintaining the focus of attention, which various studies have suggested to be, at least in part, supramodal (e.g., Eimer & Schröger, 1998), and the processes related to nonspatial feature selection, or attentional resolving, which other studies have suggested involve modality-specific sensory brain areas (Talsma & Kok, 2001).

An alternative explanation that one might consider would be that participants were distributing their visual attention somewhat differently between the upper and lower visual field positions among the four attention conditions. Whereas in the attend-RSVP condition, attention is likely to have been allocated fully to the upper visual field, in the attend-visual and attend-audiovisual conditions, visuo-spatial attention is likely to have been fully focused on the lower visual field. In the attend-auditory condition, there would be a possibility that attention was distributed somewhat in between these two locations. Such a differential distribution across visual space might result—at least in part—in a pattern of results similar to that we observed here: smaller SSVEP amplitudes in the attend-visual and attend-audiovisual conditions, large SSVEP amplitudes in the attend-RSVP condition, and somewhat intermediate amplitudes in the attend-auditory condition, because the spotlight of visual attention would still fall partly on the location of the RSVP stream.

We believe, however, that this is unlikely to be the case. First, both the visual and the auditory objects were presented from a lower hemifield position, from matched locations. As discussed above, several studies have shown that the visual and auditory attention tends to be directed to the same location in space (e.g., Eimer & Schröger, 1998; Talsma & Kok, 2002). Furthermore, previous studies related to the ventriloquism effects have shown that the apparent location of sounds can be perceptually shifted toward the location of a concurrently presented visual stimulus (Bertelson, Vroomen, De Gelder, & Driver, 2000; Vroomen, Bertelson, & De Gelder, 2001).

Another alternative explanation would be that a subjective difference in task difficulty could be responsible for the observed pattern of results. For instance, if the visual targets were harder to detect, one could argue that subjects would allocate more attention to the objects stream (and hence less to the RSVP stream) when visual objects were attended than when the auditory objects were attended. However, if the targets of one modality were indeed harder to detect than those in the other modality, we would also have expected to find significant differences in both response times and accuracy measures between the attend-visual and attend-auditory conditions. Because we did not find evidence showing that response patterns to the unisensory visual and auditory targets differed significantly from each other, we conclude that our target difficulty manipulation was successful and that we can rule out that a difference in target difficulty was responsible for generating the SSVEP effects observed in the present study.

More importantly, however, even if a differential distribution of spatial attention across the upper and lower hemispaces, or a difference in difficulty, were to contribute to the observed effects, another important finding of the present study still conflicts with the notion of a central amodal limit in attentional resources. If attention had a central capacity limitation, we would predict that the SSVEP amplitude in the attend-audiovisual condition would be the smallest, in particular, that it would be smaller than the SSVEP amplitude in the attend-visual condition alone, because the added auditory attentional demands would further reduce the capacity for processing the RSVP letters. The finding that this is not the case thus clearly argues against a single, central, capacity limit. The latter finding is also interesting in that it stands somewhat in contrast with the behavioral data. The latter show a decrease in performance, both in response time and in accuracy, in the attend-audiovisual condition. One should consider, however, that the SSVEP data is mainly representing perceptual processing, whereas the behavioral data represent the culmination of all processes, from perception to motor responses. Thus, although the present data do not imply that there is no attentional bottleneck at all in monitoring multiple sensory modalities, they do imply that this bottleneck occurs past the point of perceptual processing, taking place presumably at the higher-order cognitive processing stages instead.

In summary, we conclude that attentional modulation of sensory neural processing in visual cortex can occur at least partially independently from similar attentional modulations to auditory processing, perhaps in part to facilitate the integration of auditory and visual stimuli. This conclusion is based on the observation that visual steady-state evoked potentials differed in amplitude when attention was directed to a concurrent stream of visual objects versus to a concurrent stream of auditory objects. Furthermore, the amplitude effects observed on the SSVEP were not found to be further reduced when attention was divided across the modalities. Thus, we conclude that attentional capacity across sensory modalities is larger than within a single modality.

REFERENCES

Arnell, K. M., & Jolicœur, P. (1996). The attentional blink across modalities. *International Journal of Psychology*, *31*, 2094.

Arnell, K. M., & Jolicœur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology—Human Perception and Performance*, *25*, 630–648.

Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., et al. (2000). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience*, *12*, 988–1000.

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*, 321–332.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*, 292–297.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*, 215–229.

de Ruiter, M., Kok, A., & van der Schoot, M. (1998). Effects of inter- and intramodal selective attention to nonspatial visual stimuli. *Biological Psychology*, *49*, 269–294.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*, 808–810.

Eimer, M., & Schröger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology*, *35*, 313–327.

Eimer, M., & van Velzen, J. (2002). Crossmodal links in spatial attention are mediated by supramodal control processes: Evidence from event-related potentials. *Psychophysiology*, *39*, 437–449.

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro-Otology*, *5*, 151–166.

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and non-spatial attention. *Neuroimage*, *19*, 496–512.

Heslenfeld, D. J., Kenemans, J. L., Kok, A., & Molenaar, P. C. M. (1997). Feature processing and attention in the human visual system: An overview. *Biological Psychology*, *45*, 183–215.

Hillyard, S. A., Simpson, G. V., Woods, D. L., Van Voorhis, S., & Münte, T. F. (1984). Event-related brain potentials and selective attention to different modalities. In F. Renoso-Suarez & C. Ajmone-Marsan (Eds.), *Cortical integration* (pp. 395–414). New York: Raven Press.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, *3*, 284–291.

Jolicœur, P. (1999). Restricted attentional capacity between sensory modalities. *Psychonomic Bulletin & Review*, *6*, 87–92.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341.

LaBerge, D. (1995). *Attentional processing: The brain's art of mindfulness*. Cambridge, MA: Harvard University Press.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*, 909–919.

Müller, M. M., & Hillyard, S. (2000). Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. *Clinical Neurophysiology*, *111*, 1544–1552.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*, 309–312.

Müller, M. M., Picton, T. W., Valdes-Sosa, P., Riera, J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (1998). Effects of spatial selective attention on the steady-state visual evoked potential in the 20–28 Hz range. *Cognitive Brain Research*, *6*, 249–261.

Müller, M. M., Teder-Sälejärvi, W., & Hillyard, S. A. (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neuroscience*, *1*, 631–634.

Schröger, E., Giard, M. H., & Wolff, C. (2000). Auditory distraction: Event-related potential and behavioral indices. *Clinical Neurophysiology*, *111*, 1450–1460.

Senkowski, D., Talsma, D., Herrmann, C. S., & Woldorff, M. G. (2005). Multisensory processing and oscillatory gamma responses: Effects of selective spatial attention. *Experimental Brain Research*, *166*, 411–426.

Shomstein, S., & Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *Journal of Neuroscience*, *26*, 435–439.

Talsma, D., Doty, T. J., & Woldorff, M. G. (in press). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*. Available at: http://cercor.oxfordjournals.org/papbyrecent.dtl.

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, M. G. (2005a). Attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*, 1098–1114.

Talsma, D., & Woldorff, M. G. (2005b). Methods for the estimation and removal of artifacts and overlap in ERP waveforms. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 115–148). Cambridge, MA: MIT Press.

Vroomen, J., Bertelson, P., & DeGelder, B. (2001). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica*, *108*, 21–33.

Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1993). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, *28*, 30–42.

Woldorff, M. G., Hazlett, C. J., Fichtenholtz, H. M., Weissman, D. H., Dale, A. M., & Song, A. W. (2004). Functional parcellation of attentional control regions of the brain. *Journal of Cognitive Neuroscience*, *16*, 149–165.

(RECEIVED January 31, 2006; ACCEPTED June 23, 2006)