Behavioral/Systems/Cognitive

Attention Distributed across Sensory Modalities Enhances Perceptual Performance

Jyoti Mishra and Adam Gazzaley

Department of Neurology, Physiology and Psychiatry, Keck Center for Integrative Neurosciences, University of California, San Francisco, San Francisco, California 94158

This study investigated the interaction between top-down attentional control and multisensory processing in humans. Using semantically congruent and incongruent audiovisual stimulus streams, we found target detection to be consistently improved in the setting of distributed audiovisual attention versus focused visual attention. This performance benefit was manifested as faster reaction times for congruent audiovisual stimuli and as accuracy improvements for incongruent stimuli, resulting in a resolution of stimulus interference. Electrophysiological recordings revealed that these behavioral enhancements were associated with reduced neural processing of both auditory and visual components of the audiovisual stimuli under distributed versus focused visual attention. These neural changes were observed at early processing latencies, within 100–300 ms poststimulus onset, and localized to auditory, visual, and polysensory temporal cortices. These results highlight a novel neural mechanism for top-down driven performance benefits via enhanced efficacy of sensory neural processing during distributed audiovisual attention relative to focused visual attention.

Introduction

Our natural environment is multisensory, and accordingly we frequently process auditory and visual sensory inputs simultaneously. The neural integration of multisensory information is in turn intimately linked to the allocated focus or distribution of attention, which allows for dynamic selection and processing of sensory signals that are relevant for behavior.

Within-sensory-modality attention has been extensively characterized in the visual and auditory domains. As per the biased competition model, selective attention to a spatial location, object, or perceptual feature within a modality amplifies the sensory neural responses for the selected signal and suppresses irrelevant responses (Desimone and Duncan, 1995; Desimone, 1998; Kastner and Ungerleider, 2001; Gazzaley et al., 2005; Beck and Kastner, 2009). This mechanism has been shown to underlie improved behavioral performance for detection of the attended item. In contrast to selective attention, divided attention within a modality to concurrent sensory signals generates reduced neural responses and relatively compromised performance for each item competing for attention. These observations are mainly attributed to limited attentional resources (Lavie, 2005). A crucial question that arises is how these principles of attention extend to interactions across sensory modalities (Talsma et al., 2010).

In this study, we compare the influence of attention focused on one sensory modality (visual) to attention distributed across modalities (auditory and visual) on target detection of concurrently presented auditory and visual stimuli. Additionally, we assess neural mechanisms underlying how attention impacts performance, as well as how these behavioral and neural influences are modulated as a function of congruent versus incongruent information in the audiovisual domains. To the best of our knowledge, only two prior studies have neurobehaviorally assessed the interaction of multisensory processing and unisensory versus multisensory attention goals (Degerman et al., 2007; Talsma et al., 2007), and no study has explored these influences in the context of audiovisual stimulus congruity. Using arbitrary audiovisual stimulus combinations (simple shapes and tones), both prior studies found neural evidence for enhanced processing during multisensory attention, but did not find these neural effects to benefit perceptual performance. We aimed to resolve these inconsistencies between behavior and underlying neurophysiology in the current study via the use of a novel paradigm using inherently congruent and incongruent stimuli that are often seen and heard in the real world.

We hypothesized that distributing attention across sensory modalities, relative to focused attention to a single modality, would have differential effects for congruent versus incongruent stimulus streams. For congruent stimuli, perceptual performance could be facilitated by distributed audiovisual attention compared with focused unisensory attention. In the case of incongruent stimuli, however, distributed attention could generate greater interference and hence degrade performance, as also hypothesized in a prior behavioral study (Mozolic et al., 2008). Our be-

Received Feb. 20, 2012; revised July 20, 2012; accepted July 21, 2012.

Author contributions: J.M. and A.G. designed research; J.M. performed research; J.M. analyzed data; J.M. and A.G. wrote the paper.

This work was supported by the National Institutes of Health Grant 5R01AG030395 (A.G.) and the Program for Breakthrough Biomedical Research Grant (J.M.). We thank Jacqueline Boccanfuso, Joe Darin, and Pin-wei Chen for their assistance with data collection.

Correspondence should be addressed to either of the following: Jyoti Mishra, University of California, San Francisco-Mission Bay, Neuroscience Research Building Room 502, MC 0444 675 Nelson Rising Lane, San Francisco, CA 94158. E-mail: jyoti@gazzaleylab.ucsf.edu; or Adam Gazzaley, University of California, San Francisco-Mission Bay, Neuroscience Research Building Room 511C, 675 Nelson Rising Lane, San Francisco, CA 94158. E-mail: adam.gazzaley@ucsf.edu.

DOI:10.1523/JNEUROSCI.0867-12.2012

 0270-6474/12/3212294-09$15.00/0

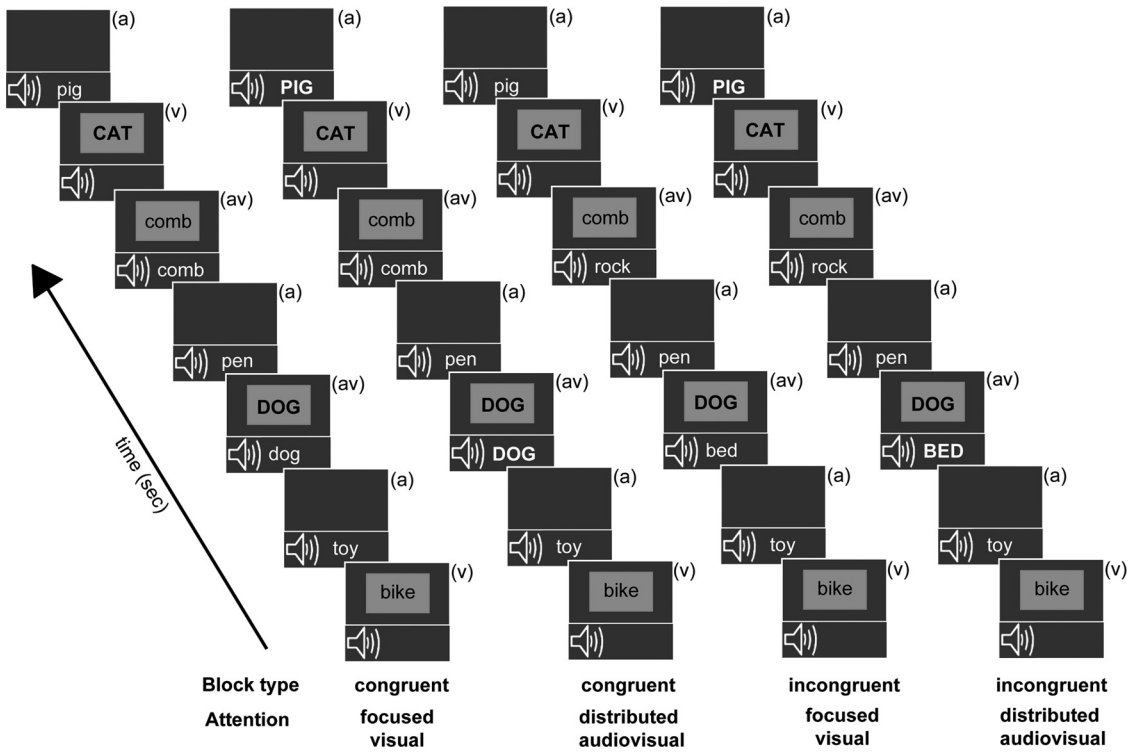

Figure 1. Overview of experimental block design. All blocks consisted of randomly interspersed auditory-only (a), visual-only (v), and simultaneous audiovisual (av) stimuli, labeled in each frame. The auditory and visual constituent stimuli of audiovisual trials matched during the two congruent blocks and did not match on incongruent blocks. Target stimuli (animal words) in each block stream are depicted in uppercase (though they did not differ in actual salience during the experiment). During the focused visual attention blocks, participants detected visual animal word targets occurring in either the V or AV stream. During the distributed audiovisual attention blocks, participants detected animal targets occurring in either of three stimulus streams.

havioral and neurophysiological results are evaluated from the perspective of these hypotheses.

Materials and Methods

Participants. Twenty healthy young adults (mean age, 23.4 years; range, 19–29 years; 10 females) gave informed consent to participate in the study approved by the Committee on Human Research at the University of California in San Francisco. All participants had normal or corrected-to-normal vision as examined using a Snellen chart and normal hearing as estimated by an audiometry software application (UHear). Additionally, all participants were required to have a minimum of 12 years of education.

Stimuli and experimental procedure. Stimuli were presented on Presentation software (Neurobehavioral Systems) run on a Dell Optiplex GX620 with a 22" Mitsubishi Diamond Pro 2040U CRT monitor. Participants were seated with a chin rest in a dark room 80 cm from the monitor. Visual stimuli (V) were words presented as black text in Arial font in a gray square sized 4.8° at the fovea. Auditory words (A) were spoken in a male voice, normalized and equated in average power spectral density, and presented to participants at a comfortable sound level of 65 dB SPL using insert earphones (Cortech Solutions). Before the experiment, participants were presented with all auditory stimuli once, which they repeated to ensure 100% word recognition. All spoken and printed word nouns were simple, mostly monosyllabic everyday usage words, e.g., tree, rock, vase, bike, tile, book, plate, soda, ice, boat, etc. The experiment used 116 unique written and corresponding spoken words; of these, 46 words were animal names (cat, chimp, cow, deer, bear, hippo, dog, rat, toad, fish, etc.) and served as targets. Visual stimuli were presented for a duration of 100 ms, all auditory presentations had a 250 ms duration, and audiovisual stimuli (AV) had simultaneous onset of the auditory and visual stimulus constituents. Each experimental run consisted of 360 randomized stimuli (shuffled from the set of 116 unique stimuli), with an equivalent 120 V alone, A alone, and AV stimulus presentations. The interstimulus interval for all stimulus types was jittered at 800–1100 ms. Each experimental block run thus lasted 6 min, with a few seconds of a self-paced break available to participants every quarter block. Stimuli were randomized at each block quarter to ensure equivalent distribution of A, V and AV stimuli in each quarter.

There were four unique block types presented randomly (Fig. 1), with each block type repeated twice and the repeat presentation occurring after each block type had been presented at least once. Participants were briefed on the upcoming block type before each block presentation: block type 1: congruent focused visual; block type 2: congruent distributed audiovisual; block type 3: incongruent focused visual; block type 4: incongruent distributed audiovisual. Block type 1 had congruent AV stimuli and participants were instructed to focus attention only on the visual stream and respond with a button press to visual animal targets, whether appearing as V alone or AV stimuli (congruent focused visual attention block). In block type 2, AV stimuli were again congruent and participants were instructed to distribute attention across both auditory and visual modalities and detect all animal names, appearing either in the V, A, or AV stream (congruent distributed audiovisual attention block). In block type 3, AV stimuli were incongruent and participants were instructed to focus attention on the visual stream only and respond to visual animal targets, either appearing alone or co-occurring with a conflicting nonanimal auditory stimulus (incongruent focused visual attention block). Lastly, in block type 4, AV stimuli were incongruent and participants distributed attention to both A and V stimuli detecting animal names in either V, A, or incongruent AV streams (incongruent distributed audiovisual attention block). Note that focused auditory block types were not included in the experiment to constrain the number of experimental manipulations and provide high-quality neurobehavioral data minimally contaminated by fatigue effects.

Targets in the A, V, or AV streams appeared at 20% probability. To further clarify, for the AV stream in congruent blocks (1 and 2), visual animal targets were paired with related auditory animal targets, while in incongruent blocks (3 and 4), visual animal targets were paired with auditory nonanimal stimuli. This AV stimuli pairing scheme was un-

known to participants and maintained the same number of visual constituent targets within the AV streams across all blocks. Note that performance metrics were obtained for targets in the V and AV streams in all blocks, while performance on targets in the A stream was only obtained in the distributed audiovisual attention blocks 2 and 4; targets in the A stream in the focused visual attention blocks 1 and 3 were not attended to and did not have associated responses.

Participants were instructed to fixate at the center of the screen at all times, and were given feedback as per their average percentage correct accuracy and response times (RTs) at the end of each block. Speed and accuracy were both emphasized in the behavior, and correct responses were scored within a 200–1200 ms period after stimulus onset. Correct responses to targets were categorized as hits; responses to nontarget stimuli in either modality were classified as false alarms. The hit and false alarm rates were used to derive the sensitivity estimate d' in each modality (MacMillan and Creelman, 1991).

EEG data acquisition. Data were recorded during eight blocks (two per block type), yielding 192 epochs of data for each standard V/A/AV stimulus (and 48 epochs per target) per block type. Electrophysiological signals were recorded with a BioSemi ActiveTwo 64-channel EEG acquisition system in conjunction with BioSemi ActiView software (Cortech Solutions). Signals were amplified and digitized at 1024 Hz with a 24-bit resolution. All electrode offsets were maintained between $\pm$20 mV.

The three-dimensional coordinates of each electrode and of three fiducial landmarks (the left and right preauricular points and the nasion) were determined by means of a BrainSight (Rogue Research) spatial digitizer. The mean Cartesian coordinates for each site were averaged across all subjects and used for topographic mapping and source localization procedures.

Data analysis. Raw EEG data were digitally re-referenced off-line to the average of the left and right mastoids. Eye artifacts were removed through independent component analyses by excluding components consistent with topographies for blinks and eye movements and the electrooculogram time series. Data were high-pass filtered at 0.1 Hz to exclude ultraslow DC drifts. This preprocessing was conducted in the Matlab (Mathworks) EEGLab toolbox (Swartz Center for Computational Neuroscience, UC San Diego). Further data analyses were performed using custom ERPSS software (Event-Related Potential Software System; UC San Diego). All ERP analyses were confined to the standard (nontarget) V, A, and AV stimuli. Signals were averaged in 500 ms epochs with a 100 ms prestimulus interval. The averages were digitally low-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 46 Hz) to remove high-frequency noise produced by muscle movements and external electrical sources. Epochs that exceeded a voltage threshold of $\pm$75 μV were rejected.

Components of interest were quantified in the 0–300 ms ERPs over distinct electrode sets that corresponded to sites at which component peak amplitudes were maximal. Components in the auditory N1 (110–120 ms) and P2 (175–225 ms) latency were measured at nine frontocentral electrodes (FC1/2, C1/2, CP1/2, FCz, Cz, CPz). Relevant early visual processing was quantified over occipital sites corresponding to the peak topography of the visual P1 component (PO3/4, PO7/8, O1/2 and POz, Oz) during the peak latency intervals of 130–140 ms and 110–130 ms for congruent and incongruent stimulus processing, respectively, and six lateral occipital electrodes (PO7/8, P7/P8, P9/P10) were used to quantify processing during the visual N1 latency (160–190 ms). Statistical analyses for ERP components as well as behavioral data used repeated-measures ANOVAs with a Greenhouse–Geisser correction when appropriate. *Post hoc* analyses consisted of two-tailed t tests. This ERP component analysis was additionally confirmed by conducting running point-wise two-tailed paired t tests at all scalp electrode sites. In this analysis, a significant difference is considered if at least 10 consecutive data points meet the 0.05 alpha criterion and is a suitable alternative to Bonferroni correction for multiple comparisons (Guthrie and Buchwald, 1991; Murray et al., 2001; Molholm et al., 2002). This analysis did not yield any new effects other than the components of interest described above.

Of note, here we refrained from analyses of later processes (>300 ms poststimulus onset), as it is not easy to distinguish whether such processes reflect a sensory/multisensory contribution or decision making/response selection processes that are active at these latencies.

Scalp distributions of select difference wave components were compared after normalizing their amplitudes before ANOVA according to the method described by McCarthy and Wood (1985). Comparisons were made over 40 electrodes spanning frontal, central, parietal, and occipital sites (16 in each hemisphere and 8 over midline). Differences in scalp distribution were reflected in significant attention condition (focused vs distributed) by electrode interactions.

Modeling of ERP sources. Inverse source modeling was performed to estimate the intracranial generators of the components within the grand-averaged difference waves that represented significant modulations in congruent and incongruent multisensory processing. Source locations were estimated by distributed linear inverse solutions based on a local auto-regressive average (LAURA) (Grave de Peralta Menendez et al., 2001). LAURA estimates three-dimensional current density distributions using a realistic head model with a solution space of 4024 nodes equally distributed within the gray matter of the average template brain of the Montreal Neurological Institute (MNI). It makes no a priori assumptions regarding the number of sources or their locations and can deal with multiple simultaneously active sources (Michel et al., 2001). LAURA analyses were implemented using CARTOOL software by Denis Brunet (http://sites.google.com/site/fbmlab/cartool). To ascertain the anatomical brain regions giving rise to the difference wave components, the current source distributions estimated by LAURA were transformed into the standardized MNI coordinate system using SPM5 software (Wellcome Department of Imaging Neuroscience, London, England).

Results

Our paradigm consisted of rapidly presented spoken (A) and written (V) nouns, either presented independently or concurrently (AV; Fig. 1). Concurrent stimuli were presented in blocks in which they were either semantically congruent (e.g., A = comb; V = comb) or incongruent (e.g., A = rock; V = comb). For both congruent and incongruent blocks, two attention manipulations were assessed: focused visual attention and distributed audiovisual attention. The participant's goal was to respond with a button-press to the presentation of a stimulus from a specific category target (i.e., animal names) when detected exclusively in the visual modality (focused attention condition) or in either auditory or visual modality (distributed attention condition). Importantly, the goals were never divided across different tasks (e.g., monitoring stimuli from multiple categories, such as animals and vehicles); thus, we investigated selective attention toward a single task goal focused within or distributed across sensory modalities. To summarize, for both congruent and incongruent blocks, two attentional variations (focused vs distributed) were investigated under identical stimulus presentations, providing the opportunity to observe the impact of top-down goals on processing identical bottom-up inputs.

Behavioral performance

Detection performance is represented by sensitivity estimates (d') and by RTs (ms) for V, A, and AV target stimuli (Table 1). d' estimates were calculated in each modality from the hits and false alarm rates for target and nontarget stimuli in that modality, respectively (MacMillan and Creelman, 1991). To compare the impact of focused versus distributed attention on multisensory processing, performance indices were generated for the difference in performance between multisensory AV and unisensory V stimuli and compared across the attentional manipulations separately for the congruent and incongruent blocks. Figure 2 shows differential (AV − V) accuracy (d') and RT metrics for distributed attention trials relative to focused attention trials in all study participants; the unity line references equivalent performance

Table 1. Details of behavioral measures observed for target stimuli during the four blocked tasks

Block type/attention	Target stimulus	*d′* (SEM)	Target hits (in %) (SEM)	Nontarget false alarms (in %) (SEM)	Reaction time (in ms) (SEM)
Congruent focused	V	5.2 (0.2)	97.5 (0.7)	0.5 (0.1)	554 (9)
	AV	5.7 (0.2)	98.3 (0.8)	0.5 (0.1)	545 (9)
Congruent distributed	V	5.0 (0.2)	97.5 (0.6)	0.8 (0.1)	548 (7)
	AV	5.9 (0.2)	99.5 (0.3)	0.9 (0.2)	523 (8)
	A	4.1 (0.2)	90.1 (1.8)	0.5 (0.1)	680 (12)
Incongruent focused	V	5.4 (0.2)	97.3 (1.2)	0.6 (0.1)	548 (9)
	AV	4.9 (0.2)	96.9 (0.7)	0.7 (0.1)	550 (8)
Incongruent distributed	V	5.1 (0.2)	97.6 (0.7)	1.0 (0.3)	538 (8)
	AV	5.4 (0.2)	98.5 (0.5)	0.9 (0.2)	544 (9)
	A	4.4 (0.2)	91.7 (2.0)	0.5 (0.1)	681 (11)

Values represented as means ± SEM.

across the two attention manipulations and the square data point represents the sample mean. Of note, there is no parallel (AV − A) performance comparison across the two attentional manipulations, as auditory targets were detected only in blocks with attention distributed to both auditory and visual inputs.

Effects of attention on congruent multisensory performance

For congruent blocks, accuracy for AV targets were significantly greater than V targets independent of attentional goals (Fig. 2*a*, positive AV − V indices), observed as a main effect of stimulus type in repeated-measures ANOVAs with stimulus type (AV vs V) and attention (focused vs distributed) as factors ($F_{(1,19)} = 20.69$, $p = 0.0002$). *Post hoc* paired *t* tests showed that AV accuracies were consistently superior to V accuracies in the focused ($t_{(19)} = 2.99$, $p = 0.007$) and the distributed attention condition ($t_{(19)} = 4.66$, $p = 0.0002$) (Fig. 2*a*, asterisks above the mean data point). This result revealed a stimulus congruency facilitation effect. There was no significant main effect of attention ($F_{(1,19)} = 0.04$, $p = 0.8$) and the interaction between the attention manipulation and stimulus type trended to significance ($F_{(1,19)} = 2.89$, $p = 0.1$). A similar ANOVA conducted for target RTs showed a comparable facilitation effect, with responses for AV targets significantly faster than V targets (main effect of stimulus type: $F_{(1,19)} = 39.50$, $p < 0.0001$; Fig. 2*b*, negative AV − V indices). Again, *post hoc* paired *t* tests showed this effect to be significant during focused ($t_{(19)} = 2.35$, $p = 0.03$) and distributed ($t_{(19)} = 7.65$, $p < 0.0001$) attention. The ANOVA for RTs additionally showed a main effect of attention ($F_{(1,19)} = 16.39$, $p = 0.0007$). Critically, a stimulus type × attention interaction was found ($F_{(1,19)} = 14.92$, $p = 0.001$), such that AV − V RTs were relatively faster in the distributed versus focused attention condition (emphasized in Fig. 2*b* by the longer *y*-axis relative to *x*-axis distance of the sample mean data point). Of note, these relatively faster AV − V RTs during distributed attention were not accompanied by any decrements in accuracy, i.e., there was no speed–accuracy tradeoff. Thus, congruent audiovisual stimuli resulted in overall better detection performance compared with visual stimuli alone (both accuracy and RT), and distributed audiovisual attention enhanced this stimulus congruency facilitation by improving performance (RT) relative to focused visual attention.

Effects of attention on incongruent multisensory performance

Similar repeated-measures ANOVAs as for congruent blocks were conducted for incongruent blocks. These revealed no main effects of stimulus type (AV vs V, $F_{(1,19)} = 0.20$, $p = 0.7$) or attention (focused vs distributed, $F_{(1,19)} = 0.23$, $p = 0.6$) on accuracy *d′* measures. Yet a significant attention × stimulus type interaction was observed ($F_{(1,19)} = 5.04$, $p = 0.04$; emphasized in Fig. 2*c* by the shorter *y*-axis relative to *x*-axis distance of the sample mean data point). *Post hoc t* tests showed that *d′* accuracy on incongruent AV targets was significantly diminished relative to V targets during focused visual attention, revealing a stimulus incongruency interference effect ($t_{(19)} = 2.13$, $p = 0.046$; Fig. 2*c*, *x*-axis asterisk on mean data point). Notably, however, this interference effect was resolved during distributed attention, such that incongruent AV target accuracy did not differ from accuracy on V targets ($t_{(19)} = 0.99$, $p = 0.3$). Neither significant main effects of attention or stimulus type nor an interaction between these factors was observed in ANOVAs for target RTs in incongruent blocks (Fig. 2*d*). Thus, distributed attention to incongruent audiovisual stimuli resulted in improved detection performance (*d′* measure) relative to focused attention, and notably without a speed–accuracy tradeoff.

Importantly, performance on visual-alone trials that served as a baseline measure (Fig. 2, horizontal zero line) did not differ as a function of condition, as evaluated in a repeated-measures ANOVA with block type (congruent vs incongruent) and attention (focused vs distributed) as factors. The main effect of block type did not reach significance for either visual accuracy ($F_{(1,19)} = 0.83$, $p = 0.4$) or RT measures ($F_{(1,19)} = 3.24$, $p = 0.09$). Similarly there was no main effect of type of attention on visual performance alone (accuracy: $F_{(1,19)} = 1.41$, $p = 0.3$; RT: $F_{(1,19)} = 2.84$, $p = 0.1$). Lastly, performance on auditory-alone targets, which only occurred in the distributed attention conditions, did not significantly differ in *d′* or RT measures across congruent versus incongruent block types.

ERP responses

Effects of attention on congruent multisensory processing

Behaviorally, we found that distributed audiovisual attention improved detection performance relative to focused visual attention for congruent audiovisual stimuli via more rapid RTs (Fig. 2*b*). As both manipulations incorporated attention to the visual modality, we investigated whether the visual constituent of the congruent AV stimulus was differentially processed under distributed versus focused attention. Visual constituent processing was obtained at occipital sites by subtracting the auditory-alone ERP from the audiovisual ERP within each attention block (Calvert et al., 2004; Molholm et al., 2004). An ANOVA with attention type as a factor conducted on the AV − A difference waves revealed significantly reduced signal amplitudes at a latency of 130–140 ms in the distributed relative to focused attention condition ($F_{(1,19)} = 4.65$, $p = 0.04$). A similar effect of attention was observed at the 160–190 ms latency range at more lateral occipital sites [$F_{(1,19)} = 5.26$, $p = 0.03$; Fig. 3, *a* (positive values plotted below horizontal axis) and *b*]. These observed AV − A differences were not driven by differences in auditory alone ERPs, which were nonsignificant across the two attention manipulations at these occipital sites.

Source estimates of the extracted visual processing signal at 130–140 and 160–190 ms modeled within the AV − A difference waves under focused attention showed neural generators in extrastriate visual cortex (in the region of BA 19; Fig. 3*c*; MNI coordinates of the peak of the source clusters in Table 2). We thus observed that these two difference wave components respectively resembled the P1 and N1 components commonly elicited in the visual-evoked potential in their timing, topography, and location of occipital source clusters (Gomez Gonzalez et al., 1994; Di Russo et al., 2002, 2003). Thus, distributed audiovisual attention was associated with reduced visual constituent processing compared with focused visual attention, which is consistent with

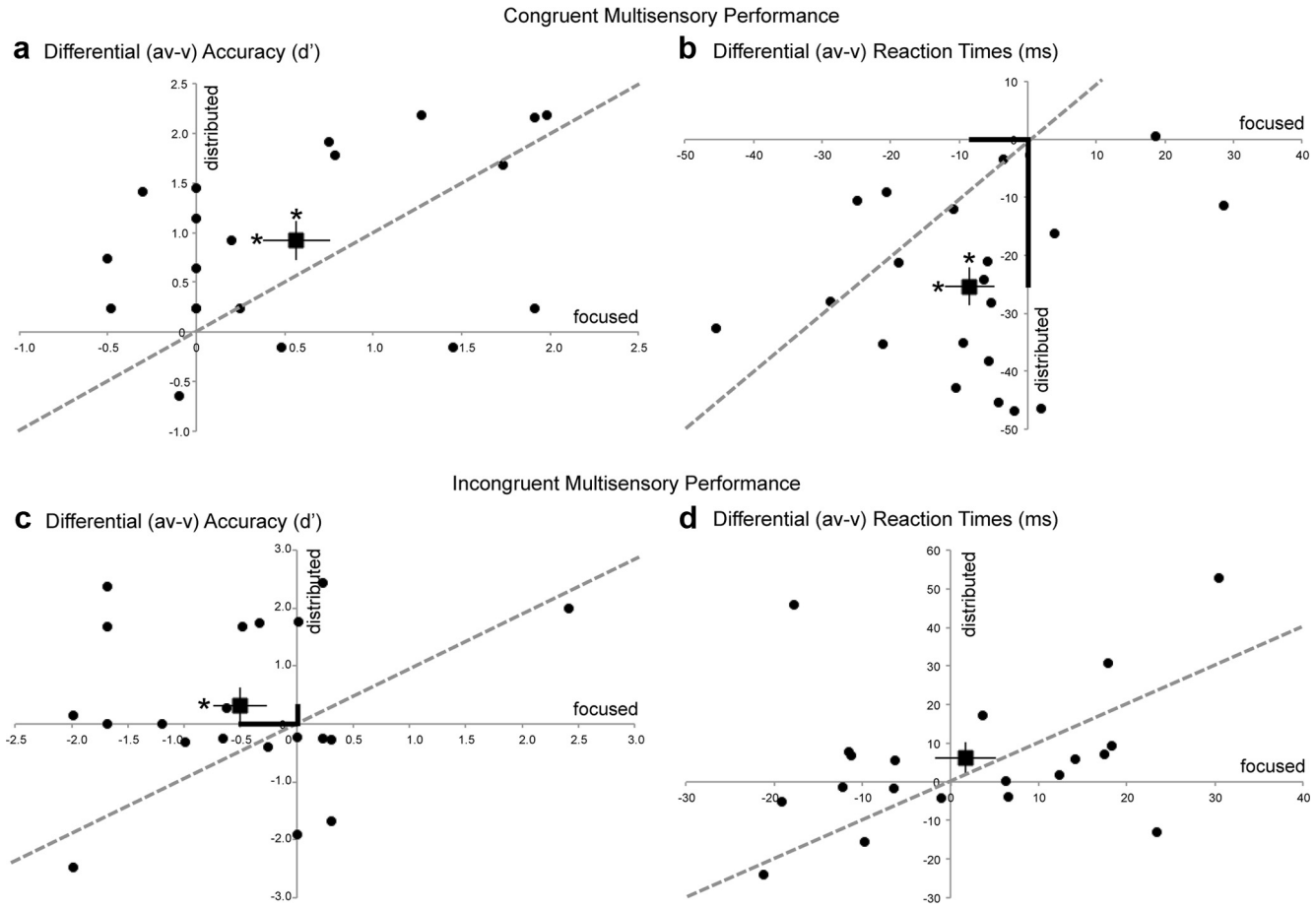

Figure 2. Behavioral performance during distributed audiovisual attention relative to focused visual attention depicted as (AV − V) normalized measures for all participants (black circles). The square data point with error bars represents the sample mean and SEM. The unity line references equivalent performance across the two attention manipulations. Measures are shown as differential *d′* (***a***, ***c***) and differential RTs (***b***, ***d***). Asterisks on the mean data points represent significant AV versus V performance differences (on horizontal error bars for focused attention and on vertical error bars for distributed attention). Bolded axial distances in ***b*** and ***c*** emphasize significant performance differences between the focused and distributed attention conditions. Note that distributed attention results in superior performance on congruent trials (RT) and incongruent trials (*d′*).

observations of sensory processing under unimodal divided attention (Desimone and Duncan, 1995; Desimone, 1998; Kastner and Ungerleider, 2001; Beck and Kastner, 2009; Reddy et al., 2009). ERPs elicited to isolated V targets under focused visual versus distributed audiovisual attention provide confirmation that unimodal visual target processing was indeed reduced in the latter condition (Fig. 3*a*, right).

In contrast to the visual modality that was attended in both focused and distributed attention conditions, the auditory modality was only attended to in the distributed condition. To compare auditory processing of the congruent AV stimulus during distributed attention versus when auditory information was task-irrelevant (i.e., during focused visual attention), we analyzed the auditory constituent at frontocentral electrode sites, where an auditory ERP is typically observed. This was accomplished by subtracting the visual-alone ERP from the audiovisual ERP for each attention condition (Calvert et al., 2004; Busse et al., 2005; Fiebelkorn et al., 2010). (The visual-alone ERPs were not different at frontocentral sites under the two types of attention.) An ANOVA with attention type as a factor conducted on the AV − V difference ERPs showed a significant early positive component difference at 175–225 ms (P200; $F_{(1,19)} = 14.3$, $p = 0.001$), which was larger when the auditory information was task-irrelevant relative to levels in the distributed attention condition (Fig. 3*d*,*e*). This difference in auditory constituent processing was positively correlated with the relative multisensory RT improvement for distributed versus focused attention observed in Figure 2*b* ($r_{18} = 0.46$, $p = 0.04$; Fig. 3*f*), revealing that reduced AV − V neural processing under distributed attention was associated with better AV − V behavioral performance.

The neural generators of the grand-averaged P200 difference wave component were modeled in the focused visual attention condition, which contained greater signal amplitude but similar scalp topography as the P200 component in the distributed condition. The source estimates revealed that the P200 component could be accounted for by bilateral current sources in the region of the superior temporal gyrus (STG; BA 22; Fig. 3*g*; MNI coordinates of the peak of the source cluster provided in Table 2). Though this P200 resembled the P2 component ubiquitously found in the auditory-evoked response (for review, see Crowley and Colrain, 2004), its localization to STG—a known site for multisensory integration (Calvert, 2001; Calvert et al., 2004; Beauchamp, 2005; Ghazanfar and Schroeder, 2006)—indicates the polysensory contribution to this process.

Effects of attention on incongruent multisensory processing

Behaviorally, we found that distributed attention improved performance relative to focused attention for incongruent audiovisual stimuli via recovery of accuracy interference costs (Fig. 2*c*). Parallel to the ERP analysis for congruent stimuli, we first ana-

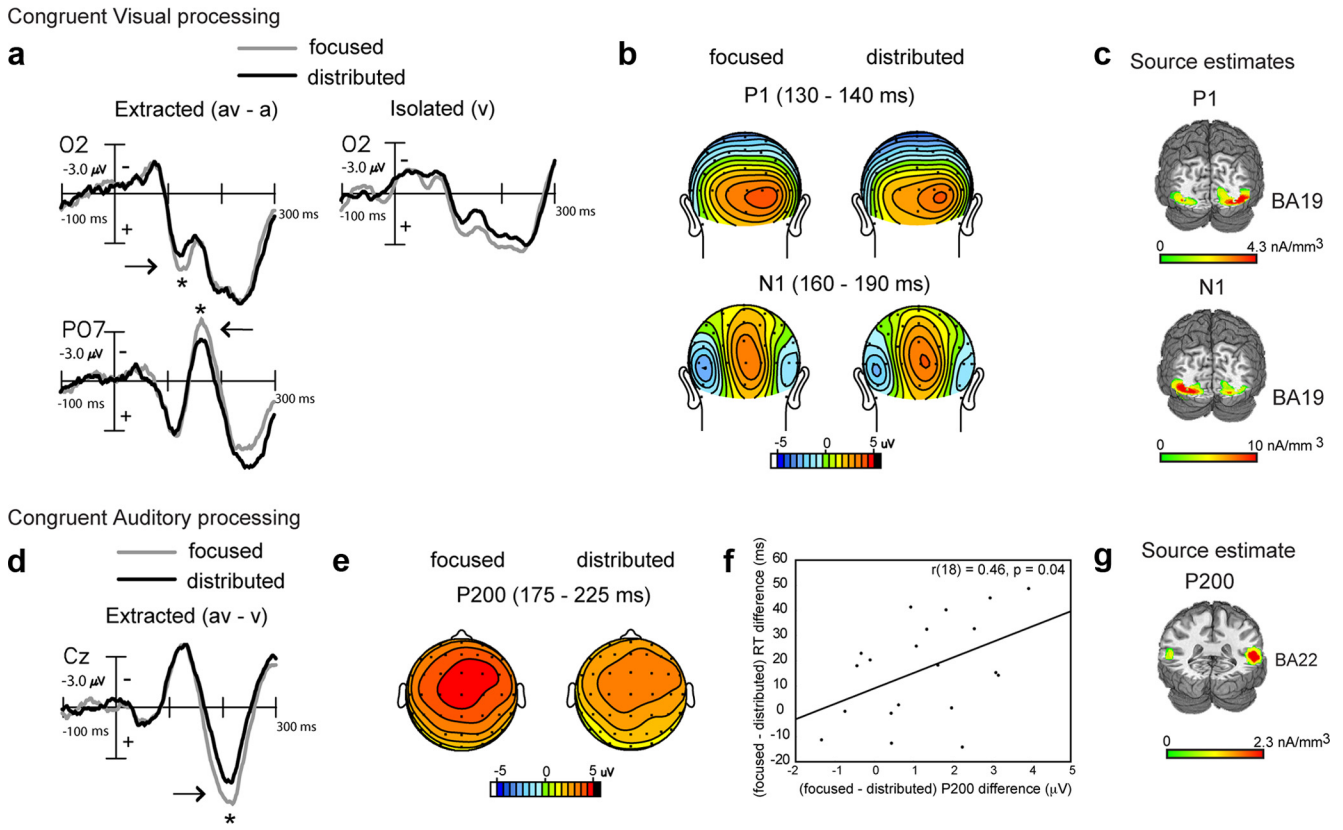

Figure 3. Grand-averaged difference waves ($n = 20$) depicting multisensory processing during the congruent trials compared for the focused and distributed attention conditions. ***a***, Extracted processing for the visual constituent of multisensory stimulation (AV − A) at occipital sites O2 and PO7 showing significant amplitude differences at 130–140 and 160–190 ms, with corresponding topographical maps in ***b*** and source estimates in ***c***. Corresponding ERPs elicited to isolated visual targets are also shown in ***a*** for reference, to the right of the extracted difference waves. ***d***, Extracted processing for the auditory constituent of multisensory stimulation (AV − V) showing attention-related differences at 175–225 ms latency (P200 component) at a medial central site (Cz; positive voltage plotted below horizontal axis). ***e***, Topographical voltage maps corresponding to the P200 component difference. ***f***, Positive neurobehavioral correlations between P200 modulation and RT differences across the two attention conditions. ***g***, Current source estimates for the P200 component.

Table 2. MNI coordinates of the peak of the source clusters as estimated in LAURA at relevant component latencies identified in the extracted visual (AV − A) and extracted auditory (AV − V) difference waveforms for congruent and incongruent blocks

Block type	Difference wave	Latency (ms)	*x* (mm)	*y* (mm)	*z* (mm)
Congruent	AV − A	130–140	±29	−71	−1
	AV − A	160–190	±27	−75	−4
	AV − V	175–225	±56	−33	+7
Incongruent	AV − A	110–120	±30	−71	−2
	AV − V	110–120	±58	−35	+4

All sources were modeled for difference waves in the focused visual attention condition.

lyzed the visual constituent of incongruent AV stimulus processing using AV − A difference waves obtained within the focused and distributed attention blocks. Early extracted visual processing signals, compared by ANOVAs with attention type as a factor, differed at occipital sites during the latency range of 110–130 ms, with significantly reduced amplitudes in the distributed relative to the focused attention AV − A difference waves ($F_{(1,19)} = 4.43$, $p = 0.04$; Fig. 4*a*,*b*); auditory-alone ERPs were no different at these sites. Source estimates of this difference wave component revealed neural generators in extrastriate visual cortex (BA 19; Fig. 4*c*; MNI coordinates in Table 2), overlapping the source estimates for the similar latency component in the congruent extracted visual difference wave. Again, early visual constituent processing of the AV stimulus was reduced under distributed relative to focused visual attention. Additionally, this result was consistent with early sensory processing of isolated V targets, which showed reduced processing under distributed audiovisual versus focused visual attention (Fig. 4*a*, right).

The extracted auditory constituent reflected in the AV − V waves in the setting of audiovisual incongruency, compared by ANOVAs with attention type as a factor, showed significant amplitude differences in early auditory processing at 110–120 ms latency ($F_{(1,19)} = 4.97$, $p = 0.04$; Fig. 4*d*,*e*); visual-alone ERPs did not significantly differ at these sites. When modeled for inverse source solutions, this difference wave component localized to the middle temporal gyrus (BA 22) adjacent to auditory cortex (BA 42; Fig. 4*f*; MNI coordinates of the peak of the source cluster in Table 2). We observed that this component resembled the auditory N1 in latency, topography, and approximate source generators. Of note, despite the overall similarity in waveforms across congruent and incongruent stimulus processing, the attention-related component differences during incongruent processing emerged at distinct latencies, and even earlier (110–120 ms) than processing differences during congruent processing (175–225 ms). Yet, consistent with results for auditory constituent processing of congruent AV stimuli, processing of the incongruent auditory constituent signal was also observed to be relatively decreased during distributed audiovisual versus focused visual attention. In this case, however, neurobehavioral correlations for the auditory N1 latency AV − V processing difference versus AV − V multisensory accuracy improvement under distributed relative to focused attention trended toward but did not reach significance ($r_{18} = 0.39$, $p = 0.09$).

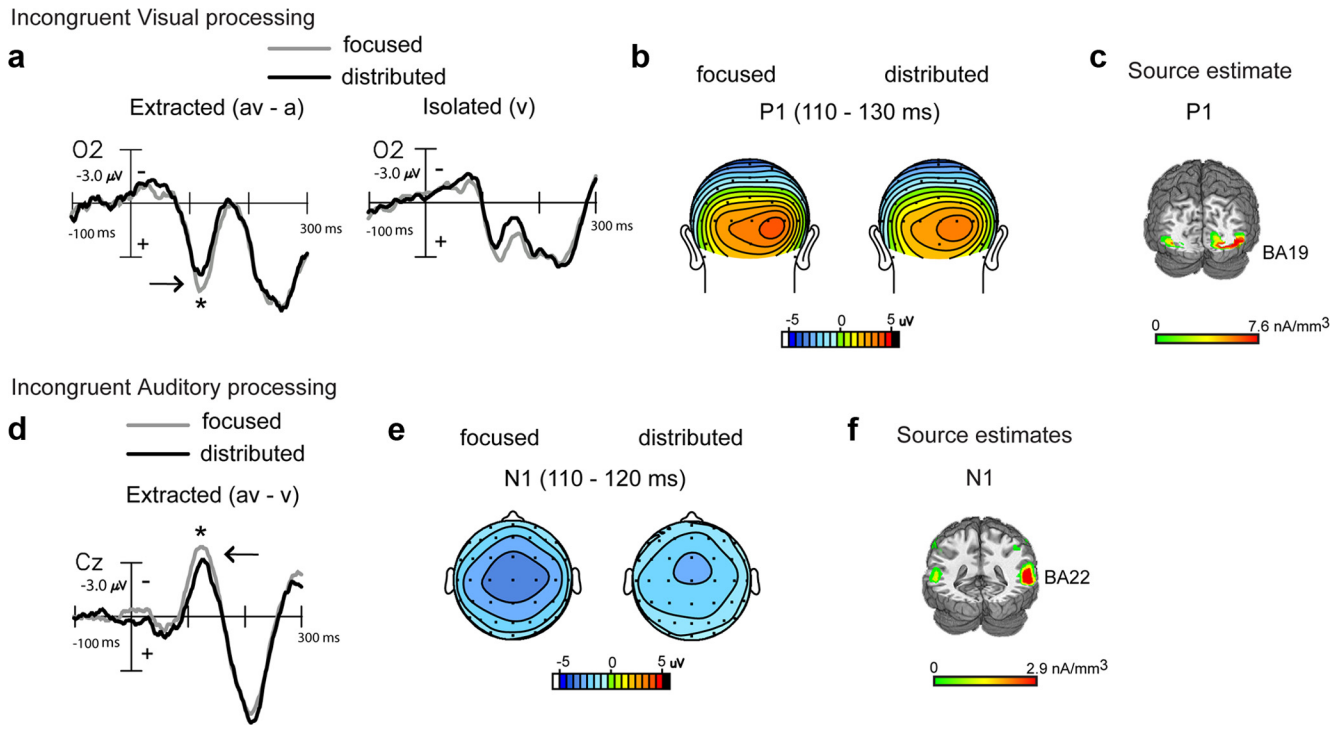

Figure 4. Grand-averaged difference waves ($n = 20$) depicting multisensory processing during the incongruent trials compared for the focused and distributed attention conditions. ***a***, Extracted processing for the visual constituent of multisensory stimulation (AV − A) at occipital site O2 showing significant amplitude differences at 110–130 ms, with corresponding topographical maps in ***b*** and source estimates in ***c***. Corresponding ERPs elicited to isolated visual targets are also shown in ***a*** for reference, to the right of the extracted difference waves. ***d***, Extracted processing for the auditory constituent of multisensory stimulation (AV − V) showing attention related differences at 110–120 ms at a medial central site (Cz), with corresponding topographical maps in ***e*** and source estimates in ***f***.

Discussion

In the present study, we investigated how processing of semantically congruent and incongruent audiovisual stimuli is influenced by the allocation of attentional focus to either a single sensory domain (visual) or distributed across the senses (auditory and visual). Behavioral findings showed that congruent audiovisual detection performance was enhanced relative to isolated visual detection during focused visual attention, and that attention distributed across both modalities further facilitated audiovisual performance via faster response times. Performance on incongruent audiovisual stimuli, in contrast, suffered a performance decrement relative to visual stimuli under focused visual attention, but, remarkably, this accuracy decrement was resolved under distributed attention. Further, event-related potential recordings consistently revealed that processing of the visual and auditory constituents of the audiovisual stimuli were markedly reduced during distributed relative to focused attention, whether or not the sensory modality in the focused condition was task-relevant (visual) or -irrelevant (auditory). Thus, these results demonstrate a novel association between improved behavioral performance and increased neural efficiency, as reflected by reduced auditory and visual processing during distributed audiovisual attention.

Previous studies using elementary auditory (tone) and visual (shape/grating) stimuli pairings have shown that perceptual performance involving multisensory stimuli is improved relative to unimodal stimuli (Giard and Peronnet, 1999; Fort et al., 2002; Molholm et al., 2002; Talsma et al., 2007; Van der Burg et al., 2011). Performance gains were consistently observed for audiovisual versus unimodal stimuli when the auditory and visual constituents of the audiovisual stimulus belonged to the same object or had prior object categorization, but not otherwise (Fort et al., 2002; Degerman et al., 2007; Talsma et al., 2007). Studies with more complex naturalistic stimuli, such as pairings of animal pictures and animal sounds (Molholm et al., 2004) and auditory speech paired with facial lip movements (Schroeder et al., 2008; Senkowski et al., 2008) have further shown that multisensory stimuli containing conflicting auditory and visual parts have negative or null performance impact relative to unimodal performance. In the majority of studies, with the exception of two (Degerman et al., 2007; Talsma et al., 2007), multisensory performance was investigated without any manipulation of the focus of attention. The two studies in exception compared performance when attention was focused unimodally or divided across an auditory and visual task; yet observations were made using arbitrary associations of elementary auditory and visual stimuli. Our study is unique in its investigation of selective attention either focused to a modality or distributed across the senses and notably for inherently congruent and incongruent stimuli.

The surprising novel behavioral finding in our study is that distributing attention across both auditory and visual domains not only enhances performance for congruent AV stimuli, but also resolves interference for incongruent AV stimuli. Although such interference resolution was unexpected, we posit that it results from efficient top-down regulation of automatic bottom-up processing when attention is distributed. During focused visual attention to incongruent AV stimuli, the concurrent and conflicting irrelevant auditory stream may capture bottom-up attention in a detrimental manner (Busse et al., 2005; Fiebelkorn et al., 2010; Zimmer et al., 2010a,b). Top-down monitoring of both sensory streams during distributed attention may minimize such interruptive bottom-up capture and may even actively suppress

the interfering stream, leading to better performance. Such a regulatory mechanism may also apply for congruent AV stimuli, wherein exclusive focus on the visual modality may weaken the automatic spatiotemporal and semantic audiovisual binding, while distributed top-down attention regulation may optimally facilitate it.

The EEG data revealed that distributed audiovisual relative to focused visual attention led to reduced neural processing for both the visual and auditory constituents of the AV stimuli. The visual constituent showed reduced ERP amplitudes during distributed attention at early visual P1 and N1 latencies; the visual P1-like attention effect was elicited for both congruent and incongruent AV processing, while the N1 effect was only observed for congruent AV processing. Congruent AV stimuli have been previously noted to generate a visual N1-like effect (Molholm et al., 2004); however, the differential modulation of early visual sensory processing by distributed versus focused attention is a novel finding. However, this is not unexpected and is consistent with the well documented finding that limited attentional resources within a modality, as during distributed relative to focused visual attention, are associated with reduced neural responses (Lavie, 2005).

Parallel to the findings for visual processing, the auditory constituent neural signals were found to be reduced during distributed attention at 200 ms peak latencies for congruent AV processing and at 115 ms peak latencies for incongruent processing. Additionally, for congruent stimuli, the amplitude reduction observed for the P200 component during distributed attention was directly correlated with the faster AV reaction times evidenced in this condition. The P200 localized to superior temporal cortex—a known site for multisensory integration (Calvert, 2001, Beauchamp, 2005)—thus, its neurobehavioral correlation underlies the polysensory contribution to the behavioral findings.

Notably, however, in the analysis for auditory constituent processing, the auditory signal under distributed audiovisual attention was compared with the signal during focused visual attention when the auditory information was task-irrelevant. The reduction in auditory constituent processing is a surprising finding given that attentional allocation is known to be associated with enhanced sensory processing. However, these neural results are consistent with the current behavioral findings and can be explained by viewing top-down attention as a dynamic regulatory process. An interfering, concurrent auditory stimulus in the case of incongruent AV stimuli has been shown to capture bottom-up attention such that auditory neural processing is enhanced (Busse et al., 2005; Fiebelkorn et al., 2010; Zimmer et al., 2010a,b). We hypothesize that distributed top-down attention reduces this bottom-up capture by the interfering auditory stream and/or may suppress the interfering stream, resulting in reduced early auditory processing and a resolution of behavioral interference effects, as observed here.

Of note, we found all neural processing differences to be amplitude rather than latency modulations. That no significant latency differences were found in our comparisons may be attributed to our focus on early multisensory processing (0–300 ms), which has been evidenced to be rapid and convergent within unisensory cortices (Schroeder and Foxe, 2002, 2005) with neural response enhancement-/suppression-related modulation mechanisms (Ghazanfar and Schroeder, 2006). Also, as a focused auditory attention manipulation was not included in the study, the current findings may be specific to comparisons of distributed audiovisual attention versus focused visual attention; modality generalization needs to be pursued in future research.

The two prior neurobehavioral studies that manipulated attention (focus unimodally or divide attention across auditory and visual tasks) found evidence for enhanced early multisensory ERP processing (Talsma et al., 2007) and enhanced multisensory-related fMRI responses in superior temporal cortex (Degerman et al., 2007) under divided attention. These enhancements were associated with null or negative multisensory performance (under divided relative to unimodal attention) in the former and latter study, respectively. Although these findings appear contrary to our results, these studies differ from the current investigation in the use of elementary auditory and visual stimuli with no inherent (in)congruencies and in the manipulation of attention being divided across distinct auditory and visual tasks in contrast to selective attention directed at a single task being distributed across modalities, as in the current study. These crucial study differences may be manifest in the different underlying neural changes and behavioral consequences. To note, consistent with these studies and contrary to ours, a prior behavioral study that used a similar task design as us did not find any significant performance differences for semantically incongruent audiovisual pairings processed under unisensory versus multisensory attention goals (Mozolic et al., 2008). This study, however, used a two-alternative forced-choice response scheme, different and more cognitively complex than the detection scheme in the present experiment. It is possible then that any interference resolution under distributed attention, as we observe, may be annulled in the Mozolic et al. (2008) study by the additional conflict introduced at the level of decision making to choose the appropriate response.

Many cross-modal studies to date have reported that multisensory attention is associated with enhancements in neural processing (for review, see Talsma et al., 2010). However, we posit that distributed audiovisual attention, which was beneficial for multisensory behavior relative to focused visual attention, is characterized by reduced sensorineural processing. Such associations between improved performance and reduced neural processing have been more commonly observed in the perceptual learning literature. Recent ERP and fMRI studies have evidenced that perceptual training associated with improved behavioral performance results in reduced neural processing in perceptual (Ding et al., 2003; Mukai et al., 2007; Alain and Snyder, 2008; Kelley and Yantis, 2010) and working memory (Berry et al., 2010) tasks, and that individuals with trained attentional expertise exhibit reduced responses to task-irrelevant information (Mishra et al., 2011). These training-induced plasticity studies interpret these data as a reflection of increased neural efficacy impacting improved behavioral performance. Overall, our findings show that distributing attention across the senses can be beneficial in a multisensory environment, and further demonstrate novel neural underpinnings for such behavioral enhancements in the form of reduced processing within unisensory auditory and visual cortices and polysensory temporal regions.

References

Alain C, Snyder JS (2008) Age-related differences in auditory evoked responses during rapid perceptual learning. Clin Neurophysiol 119:356–366.

Beauchamp MS (2005) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol 15:145–153.

Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. Vision Res 49:1154–1165.

Berry AS, Zanto TP, Clapp WC, Hardy JL, Delahunt PB, Mahncke HW,

Gazzaley A (2010) The influence of perceptual training on working memory in older adults. PLoS One 5:e11537.
Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci U S A 102:18751–18756.
Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123.
Calvert GA, Spence C, Stein BE (2004) The handbook of multisensory processing. Cambridge, MA: MIT.
Crowley KE, Colrain IM (2004) A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin Neurophysiol 115:732–744.
Degerman A, Rinne T, Pekkola J, Autti T, Jääskeläinen IP, Sams M, Alho K (2007) Human brain activity associated with audiovisual perception and attention. Neuroimage 34:1683–1691.
Desimone R (1998) Visual attention mediated by biased competition in extrastriate visual cortex. Philos Trans R Soc Lond B Biol Sci 353:1245–1255.
Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222.
Ding Y, Song Y, Fan S, Qu Z, Chen L (2003) Specificity and generalization of visual perceptual learning in humans: an event-related potential study. Neuroreport 14:587–590.
Di Russo F, Martínez A, Sereno MI, Pitzalis S, Hillyard SA (2002) Cortical sources of the early components of the visual evoked potential. Hum Brain Mapp 15:95–111.
Di Russo F, Martínez A, Hillyard SA (2003) Source analysis of event-related cortical activity during visuo-spatial attention. Cereb Cortex 13:486–499.
Fiebelkorn IC, Foxe JJ, Molholm S (2010) Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter? Cereb Cortex 20:109–120.
Fort A, Delpuech C, Pernier J, Giard MH (2002) Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res Cogn Brain Res 14:20–30.
Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M (2005) Top-down enhancement and suppression of the magnitude and speed of neural activity. J Cogn Neurosci 17:507–517.
Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278–285.
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.
Gomez Gonzalez CM, Clark VP, Fan S, Luck SJ, Hillyard SA (1994) Sources of attention-sensitive visual event-related potentials. Brain Topogr 7:41–51.
Grave de Peralta Menendez R, Gonzalez Andino S, Lantz G, Michel CM, Landis T (2001) Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. Brain Topogr 14:131–137.
Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. Psychophysiology 28:240–244.
Kastner S, Ungerleider LG (2001) The neural basis of biased competition in human visual cortex. Neuropsychologia 39:1263–1276.
Kelley TA, Yantis S (2010) Neural correlates of learning to attend. Front Hum Neurosci 4:216.
Lavie N (2005) Distracted and confused? Selective attention under load. Trends Cogn Sci 9:75–82.
MacMillan NA, Creelman CD (1991) Detection theory: a user's guide. New York: Cambridge UP.
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.
Michel CM, Thut G, Morand S, Khateb A, Pegna AJ, Grave de Peralta R, Gonzalez S, Seeck M, Landis T (2001) Electric source imaging of human brain functions. Brain Res Brain Res Rev 36:108–118.
Mishra J, Zinni M, Bavelier D, Hillyard SA (2011) Neural basis of superior performance of action videogame players in an attention-demanding task. J Neurosci 31:992–998.
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. Brain Res Cogn Brain Res 14:115–128.
Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452–465.
Mozolic JL, Hugenschmidt CE, Peiffer AM, Laurienti PJ (2008) Modality-specific selective attention attenuates multisensory integration. Exp Brain Res 184:39–52.
Mukai I, Kim D, Fukunaga M, Japee S, Marrett S, Ungerleider LG (2007) Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. J Neurosci 27:11401–11411.
Murray MM, Foxe JJ, Higgins BA, Javitt DC, Schroeder CE (2001) Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. Neuropsychologia 39:828–844.
Reddy L, Kanwisher NG, VanRullen R (2009) Attention and biased competition in multi-voxel object representations. Proc Natl Acad Sci U S A 106:21447–21452.
Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. Curr Opin Neurobiol 15:454–458.
Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res Cogn Brain Res 14:187–198.
Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 12:106–113.
Senkowski D, Saint-Amour D, Gruber T, Foxe JJ (2008) Look who's talking: the deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. Neuroimage 43:379–387.
Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex 17:679–690.
Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci 14:400–410.
Van der Burg E, Talsma D, Olivers CN, Hickey C, Theeuwes J (2011) Early multisensory interactions affect the competition among multiple visual objects. Neuroimage 55:1208–1218.
Zimmer U, Itthipanyanan S, Grent-'t-Jong T, Woldorff MG (2010a) The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention. Eur J Neurosci 31:1744–1754.
Zimmer U, Roberts KC, Harshbarger TB, Woldorff MG (2010b) Multisensory conflict modulates the spread of visual attention across a multisensory object. Neuroimage 52:606–616.