

Cross-modal Emotional Attention: Emotional Voices Modulate Early Stages of Visual Processing

Tobias Brosch, Didier Grandjean, David Sander, and Klaus R. Scherer

Abstract

Emotional attention, the boosting of the processing of emotionally relevant stimuli, has, up to now, mainly been investigated within a sensory modality, for instance, by using emotional pictures to modulate visual attention. In real-life environments, however, humans typically encounter simultaneous input to several different senses, such as vision and audition. As multiple signals entering different channels might originate from a common, emotionally relevant source, the prioritization of emotional stimuli should be able to operate across modalities. In this study, we explored cross-modal emotional attention. Spatially localized utterances with emotional and neutral prosody served as cues for a visually presented target in a cross-modal dot-probe task. Participants

were faster to respond to targets that appeared at the spatial location of emotional compared to neutral prosody. Event-related brain potentials revealed emotional modulation of early visual target processing at the level of the P1 component, with neural sources in the striate visual cortex being more active for targets that appeared at the spatial location of emotional compared to neutral prosody. These effects were not found using synthesized control sounds matched for mean fundamental frequency and amplitude envelope. These results show that emotional attention can operate across sensory modalities by boosting early sensory stages of processing, thus facilitating the multimodal assessment of emotionally relevant stimuli in the environment. ■

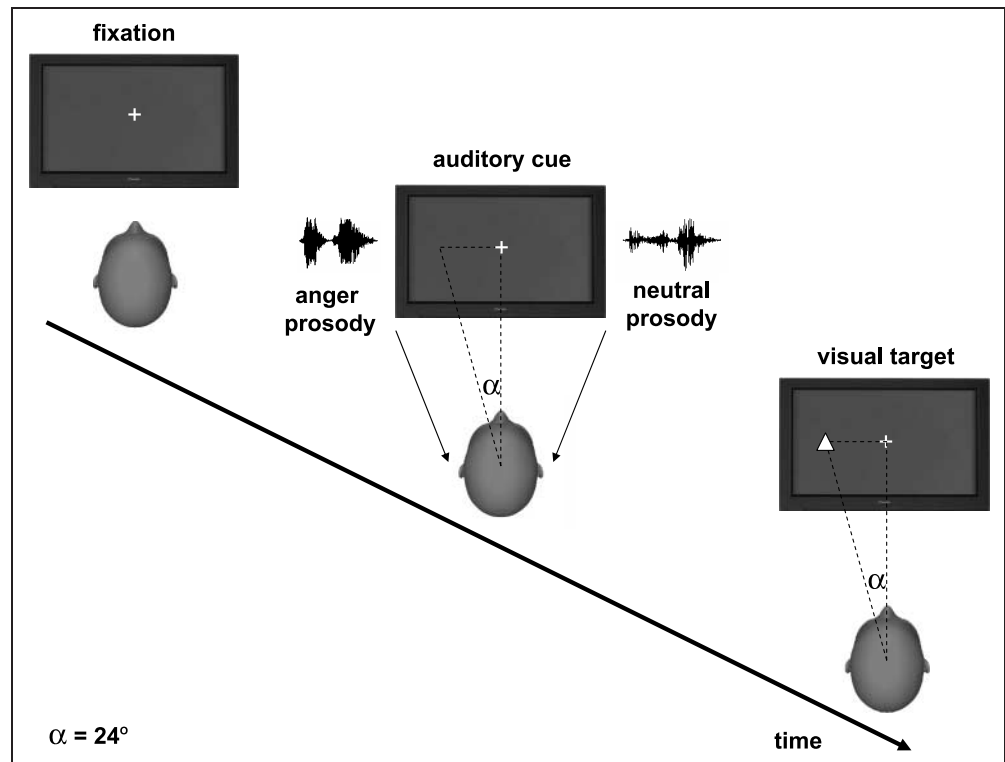
INTRODUCTION

The human organism is constantly confronted with a huge amount of stimulus input from the environment. Due to limited capacity (Marois & Ivanoff, 2005), the brain cannot exhaustively process all the input and has to select some stimuli at the cost of others (Desimone & Duncan, 1995). In addition to basic physical features such as color or size (Wolfe & Horowitz, 2004), emotional relevance is an important dimension which can modulate this process. Emotional stimuli are privileged in the competition for neural processing resources. Brain activation elicited by emotional stimuli (such as pictures, words, or sounds) is higher than for neutral stimuli, reflecting a more robust and stable neural representation (Vuilleumier, 2005; Davidson, Maxwell, & Shackman, 2004). A number of brain imaging studies have shown that detection and preferential processing of emotional stimuli occurs even when they are not initially in the focus of attention (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2006; Grandjean et al., 2005; Vuilleumier, Armony, Driver, & Dolan, 2001). The amygdala, a neural structure in the medial-temporal lobe with extensive connections to many other brain regions (LeDoux, 2000), is crucially involved in the preferential processing of emotional stimuli. For example, amygdala activity is correlated with enhanced responses to emotional stimuli in the visual cortex (Morris et al., 1998). Furthermore, amygdala lesions can abolish the enhanced activation for emotional compared to neutral faces in the visual cortex (Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004). Thus, it has been suggested that increased perceptual processing of emotional stimuli results from direct feedback signals from the amygdala to cortical sensory pathways (Vuilleumier, 2005).

The preferential treatment of emotional stimuli is reflected in participants' behavior in several cognitive paradigms, such as the visual search task (Brosch & Sharma, 2005; Öhman, Flykt, & Esteves, 2001), the attentional blink paradigm (Anderson, 2005), the attentional cueing paradigm (Fox, Russo, & Dutton, 2002), and the dot-probe task (Brosch, Sander, & Scherer, 2007; Lipp & Derakshan, 2005; Mogg & Bradley, 1999). In the dot-probe task (see Figure 1), participants respond to the location or identity of a target, which replaces one out of two simultaneously presented cues. One of the cues is emotional, the other one is neutral. Behavioral results in the dot-probe task show facilitated processing when the target replaces the emotional cue compared to the neutral cue, reflected by faster response times toward the targets (Brosch et al., 2007; Lipp & Derakshan, 2005). This is interpreted

University of Geneva, Switzerland

Figure 1. Experimental sequence.



as the result of attentional capture by the emotional stimulus, which then leads to increased processing of the target.

Event-related potentials (ERPs) recorded during the emotional dot-probe task reveal an augmentation of the P1 component elicited by a target replacing the emotional compared to the neutral cue (Brosch, Sander, Pourtois, & Scherer, 2008; Pourtois, Grandjean, Sander, & Vuilleumier, 2004). Earlier ERP results have indicated that the P1 exogenous visual response is systematically enhanced in amplitude in response to attended relative to unattended spatial locations or stimuli (e.g., Luck, Woodman, & Vogel, 2000). Amplitude modulations of the P1 as a function of the deployment of visuospatial attention is thought to reflect a sensory gain control mechanism causing increased perceptual processing in the visual cortex of attended locations or stimuli (Hillyard, Vogel, & Luck, 1998). The faster response times for targets replacing emotional cues in the dot-probe paradigm are thus associated with modulations of early perceptual processing of the target (and not due to postperceptual processes at the level of response selection or action preparation).

Enhanced sensory representations of emotional stimuli have been found not only for the visual (Vuilleumier et al., 2001; Morris et al., 1998) but also for the auditory domain. Several fMRI studies have shown that emotional prosody increases activity in the associative auditory cortex (superior temporal sulcus), more particularly in the sensitive voice regions (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). This effect was observed for positive and negative emotions (Ethofer et al., 2006), and emerged

even when the focus of voluntary attention was directed away from the emotional auditory stimuli using a dichotic listening task (Grandjean et al., 2005; Sander et al., 2005). Furthermore, stroke patients with left auditory extinction showed a detection increase of emotional compared to neutral prosody stimulation on the left side, showing that emotion is able to moderate an auditory extinction phenomenon (Grandjean, Sander, Lucas, Scherer, & Vuilleumier, 2008), as previous studies have already shown for the visual domain (Vuilleumier & Schwartz, 2001).

Until now, studies investigating emotional modulation of spatial attention have mainly examined *within-modality* effects, most frequently using pictures of emotional stimuli to modulate visual attention. However, humans typically encounter simultaneous input to several different senses, such as vision and audition. Signals entering these different channels might originate from a common source, requiring mechanisms for the integration of information (including emotional information) conveyed by multiple sensory channels. To receive maximal benefit from multimodal input, the brain must coordinate and integrate the input appropriately so that signals from a relevant common source are processed across the different input channels. This integration is a computational challenge, as the properties of the information representation differ greatly between the input channels (Driver & Spence, 1998).

The questions to which extent attention operates independently within each sensory modality and by which mechanisms attention is coordinated across modalities

have been investigated using simple nonemotional stimuli such as flashes of light or bursts of noise (Eimer & Driver, 2001; Driver & Spence, 1998). The paradigm most frequently used for the investigation of cross-modal attentional modulation is the spatial cueing paradigm (Posner, 1980). In this paradigm, participants indicate whether a target appeared either in the left or the right visual field. Before the target, a spatially nonpredictive peripheral cue in another modality is presented (e.g., an auditory cue preceding a visual target). Although the cue is not predictive of the location of the target, responses to the targets are faster and/or more accurate when the targets are presented on the same side as the cue (McDonald & Ward, 2000; Spence & Driver, 1997).

Like for its unimodal counterpart, ERP recordings have been used to examine the neural correlates of the cross-modal attentional modulation effect (Eimer & Driver, 2001; McDonald & Ward, 2000). In an ERP study of exogenous attentional cueing using auditory cues and visual targets, an attentional negativity (Nd) was elicited for visual ERPs recorded from lateral occipital sites (PO7/PO8) between 200 and 400 msec after stimulus onset for valid compared to invalid trials (McDonald & Ward, 2000). No cueing effects were observed for the P1 component. This suggests that cross-modal effects of a nonemotional auditory event on visual processes may be located at a stage after the initial perceptual processing of visual information.

Not much is known about the modulatory effect of emotional stimuli on attention across modalities. Automatic enhanced sensory responses of specific brain areas to emotional events have been shown both for visual (Vuilleumier et al., 2001) and auditory (Grandjean et al., 2005; Sander et al., 2005) events. This probably reflects a fundamental principle of human brain organization, namely to prioritize the processing of emotionally relevant stimuli, even if they are outside the focus of attention. Such a mechanism should be able to operate across modalities, as multiple signals entering different channels might originate from a common, emotionally relevant source. Consistent with this view, we recently showed that emotional prosody, the changes in the tone of the voice that convey information about a speaker's emotional state (Scherer, Johnstone, & Klasmeyer, 2003), can facilitate detection of a visual target (Brosch, Grandjean, Sander, & Scherer, 2008). In this cross-modal emotional dot-probe paradigm (see MacLeod, Mathews, & Tata, 1986), participants indicated the location of a visual target that was preceded by a binaurally presented pair of auditory pseudowords, one of which was uttered with anger prosody (in one ear), the other one with neutral prosody (in the other ear). Although delivered through headphones, the emotional and neutral auditory stimuli were spatialized to produce the compelling illusion that they originated from a distinctive source localized either in the left or right personal space (see Methods for details). Response

times toward (nonemotional) visual targets were shorter when they appeared in a position spatially congruent with the perceived source of the emotional prosody (Brosch, Grandjean, et al., 2008).

The aim of the present study was to investigate the neural underpinnings of cross-modal modulation of visual attention by emotional prosody. Of special interest was the question of whether cross-modal emotional attention affects early sensory stages of processing—as might be expected on the basis of investigations of emotional attention within one modality (Brosch, Sander, et al., 2008; Pourtois et al., 2004), or not—as might be expected on the basis of investigations of nonemotional cross-modal attention modulation (McDonald & Ward, 2000).

We recorded ERPs while participants performed the cross-modal emotional dot-probe task (Brosch, Grandjean, et al., 2008). Based upon earlier work investigating the modulation of visual attention by visual emotional stimuli (Brosch, Sander, et al., 2008; Pourtois et al., 2004), we predicted that a cross-modal emotional modulation of early sensory states would manifest as a modulation of the amplitude of the P1 component in form of larger amplitudes toward validly cued targets (see Figure 1) than toward invalidly cued targets.

METHODS

Participants

Seventeen students of the University of Geneva participated in the experiment. Data from two female participants were excluded due to poor quality of the physiological recording, leaving a final sample of 15 participants (13 women, mean age = 21.4 years, $SD = 3.3$). All participants were right-handed, had normal self-reported audition and normal or corrected-to-normal vision, and had no history of psychiatric or neurological disease.

Stimuli

The auditory stimuli consisted of meaningless but word-like utterances (pseudowords “goster,” “niuvenci,” “figotleich”) pronounced with either anger or neutral prosody. Sixty different utterances by 10 different speakers with a duration of 750 msec (50% male speakers, 50% anger prosody) were extracted from a database of pseudosentences that had been acquired and validated in earlier work (Banse & Scherer, 1996). The anger stimuli were directly adopted from the database, the neutral stimuli were selected from the “boredom” and “interest” stimuli, selecting the most neutral on the basis of a judgment study investigating the “neutrality” and “emotionality” of these stimuli. Fifteen participants (9 women, mean age = 25.3 years) judged the stimuli on two visual analog rating scales (“neutral” and “emotional”). Based

on those ratings, the 20 “interest” and 20 “boredom” stimuli with minimal “emotional” ratings and maximal “neutral” ratings were selected. Additionally, we performed a judgment study on the excerpts selected for the present experiment (anger, neutral) as well as emotional prosody excerpts not used in the current study (sadness, happiness, and fear). This was done to test the recognizability of the different emotional stimuli and to be sure that the neutral stimuli are perceived as “neutral” rather than “interest” or “boredom.”

Sixteen participants (undergraduate students, 14 women) judged on visual analog scales (from “not at all” to “totally”) to what extent the excerpts were pronounced with anger, neutral, boredom, interest, despair, elation, pride, disgust, contempt, happiness, sadness, fear, and surprise emotional intonation. A test of repeated measures ANOVA using the within-subjects factors emotional prosody and emotion scale revealed, as predicted, an interaction effect [$F(48, 912) = 75.78, p < .001$]. Anger stimuli were mainly rated as expressing “anger” [contrast “anger” scale vs. other scales: $F(1, 19) = 459.46, p < .001$] and neutral stimuli were mainly rated as “neutral” [contrast “neutral” scale vs. other scales: $F(1, 19) = 87.88, p < .001$]. A contrast comparing the “neutral,” “boredom,” and “interest” ratings for the neutral stimuli showed that the neutral stimuli were rated significantly higher on the “neutral” scale than on the “boredom” or “interest” scale [contrast neutral vs. boring–interest: $F(1, 19) = 52.94, p < .01$]. All stimuli were combined to 40 stereophonically presented paired utterances containing one angry and one neutral utterance. To avoid interactions of speaker sex and emotionality in stimulus pairs, only utterance pairs from same-sex speakers were combined. Each pair was matched for mean acoustic energy.

The fundamental frequency F_0 and the distribution of energy in time play an important role in conveying emotional information in voices (Grandjean, Bänziger, & Scherer, 2006; Banse & Scherer, 1996). In addition to these low-level stimulus properties, emotional information in prosody is conveyed by other, more complex perceived acoustical characteristics corresponding to objective acoustical parameters, such as spectral energy distribution in time or the temporal dynamic of the F_0 (see e.g., Banse & Scherer, 1996). The complex interactions of these different acoustical parameters over time are crucial for emotional prosody perception. To control for the low-level physical properties of our stimuli related to prosody, we included a control condition by synthesizing control stimuli matched for the mean fundamental frequency and the amplitude envelope of each vocal stimulus used in the experiment using Praat. After controlling for the low-level stimulus properties, any effect reflecting voice-specific processes that is not driven by a particular range of frequency or a specific amplitude contour should only be found for the prosody cues, not for the control cues.

In order to give the subjective impression that the sounds originate from a specific location in space, we manipulated the interaural time difference (ITD) of the sounds using a head-related transfer function (HRTF) implemented in the plug-in Panorama used with SoundForge (for more details about this procedure, see e.g., Spierer, Meuli, & Clarke, 2007). The audio pairs were transformed via binaural synthesis to be equivalent to sound sources at a distance of 110 cm and at an angle of 24° to the left and to the right of the participants (see Figure 1). We used spatially localized stimuli instead of the simpler dichotic presentation mode, as it is a closer approximation of real-life contexts in which concomitant auditory and visual information can originate from a common source localized in space. The HRTF method enables us to investigate the relationship between emotion and spatial attention processes based on realistic spatial localization rather than investigating ear effects. Previous studies with brain-damaged patients have shown a double dissociation between auditory extinction and ear extinction, highlighting the fact that these two processes are very different in terms of the brain regions involved (Spierer et al., 2007).

The experiment was controlled by E-Prime. The auditory cues were presented using Sony MDR-EX71 headphones. The visual targets were presented using a Sony VPL CX 10 projector.

Procedure

Figure 1 shows the experimental sequence. During the whole experiment, a fixation cross was presented. Each trial started with a random time interval between 500 and 1000 msec, after which the acoustic cue sound pair was presented. One of the sounds in the pair had emotional prosody, the other one neutral prosody.

The target, a neutral geometric figure (a triangle which could either point upward or downward), was presented with a variable cue–target stimulus onset asynchrony (SOA) of 550, 600, 650, 700, or 750 msec after sound onset. The target was presented for 100 msec on the left or right side, at a distance of 45 cm from the fixation cross. The participants were seated at 100 cm from the projection screen. Thus, the angle between the target and the fixation cross was 24° , which is equivalent to the synthesized location of the audio stimulus pairs. In a *valid* trial, the target appeared on the side of the emotional sound, whereas in an *invalid* trial, the target appeared on the side of the neutral sound. Valid and invalid trials were presented in randomized order with an equal proportion of valid and invalid trials (50%). Participants were instructed to press the “B” key of the response keyboard using the index finger of their right hand only when the orientation of the triangle corresponded to their respective GO condition (triangle pointing upward or downward, counterbalanced across participants). Participants had a maximum of 1500 msec

to respond, after that time, the next trial started. The experiment consisted of one practice block of 10 trials, followed by four experimental blocks of 160 trials each (total 640 trials). In two blocks, sounds with emotional and neutral prosody were presented, and in two blocks, the synthesized control sounds were presented. We designed a small number of go trials which required a motor response (10%) to study covert spatial orienting toward emotional stimuli in a vast majority of trials where there is no overt motor response (90% no-go trials), therefore minimizing the contamination of motor preparation or execution on EEG signal quality.

EEG Recordings

EEG was recorded with a sampling rate of 512 Hz using the ActiveTwo system (BioSemi, Amsterdam, Netherlands). Horizontal and vertical EOGs were recorded using four facial electrodes placed on the outer canthi of the eyes and in the inferior and superior areas of the left orbit. Scalp EEG was recorded from 64 Ag/AgCl electrodes attached to an electrode cap and positioned according to the extended 10–20 EEG system. The EEG electrodes were referenced off-line to average reference. The data were filtered using a high pass of 0.53 Hz and a low pass of 30 Hz. Data were downsampled to 256 Hz and segmented around target onsets in epochs of 1000 msec (from –200 msec to +800 msec). A reduction of artifacts related to vertical eye movements was implemented using the algorithm developed by Gratton, Coles, and Donchin (1983). A baseline correction was performed on the prestimulus interval using the first 200 msec. EEG epochs exceeding 70 μ V were excluded from the analysis. The artifact-free epochs were averaged separately for each electrode, condition, and individual. Grand-average ERPs were finally generated by computing the mean ERPs across participants in each condition.

Data Analysis

Behavioral Data

Response times for correct responses between 200 and 1000 msec were analyzed in a $2 \times 2 \times 2$ repeated measures ANOVA with the factors voice condition (prosody/synthesized control sounds), cue validity (valid/invalid) and target position (left/right).

EEG Experiment

Based on our a priori hypotheses and on inspection of the present ERP dataset, we analyzed the P1 component (130–190 msec) time-locked to the onset of the target in valid and invalid trials. Peak amplitudes and latencies were measured at lateral occipital sites (PO7/O1 and PO8/O2; see Figure 3). These sites were selected on the basis

of related effects in previous studies (Brosch, Sander, et al., 2008; Pourtois et al., 2004; Martinez et al., 1999) and on conspicuous topographic properties of the present ERP dataset. The amplitudes and latencies of the P1 were analyzed using $2 \times 2 \times 2 \times 2 \times 2$ ANOVAs with the repeated factors voice condition (prosody/synthesized control sounds), cue validity (valid/invalid), target position (left/right), hemisphere (left/right), and electrode position (PO/O). To estimate the likely configuration of intracranial neural sources underlying the observed scalp topographic maps of interest, we used a distributed inverse solution method on the basis of a Local Auto-Regressive Average model of the unknown current density of the brain (LAURA; see Grave de Peralta Menendez, Gonzalez Andino, Lantz, Michel, & Landis, 2001). The method is derived from biophysical laws describing electric fields in the brain. It computes a three-dimensional reconstruction of the generators of the brain's electromagnetic activity measured at the scalp on the basis of biophysically driven inverse solutions without a priori assumptions on the number and position of the possible generators (see also Michel et al., 2004, for further details).

RESULTS

Behavioral Data

Figure 2 shows the response times for valid and invalid trials in the prosody condition and the control condition.

There was a trend toward a Voice condition \times Cue validity interaction [$F(1, 14) = 2.51, p = .14$]. In the prosody condition, participants responded faster toward valid (549 msec) than toward invalid (565 msec) targets, as indicated by a marginally significant t test [$t(14) = 1.68, p = .06$, one-tailed], thus replicating our previous behavioral findings (Brosch, Grandjean, et al., 2008). Note that in

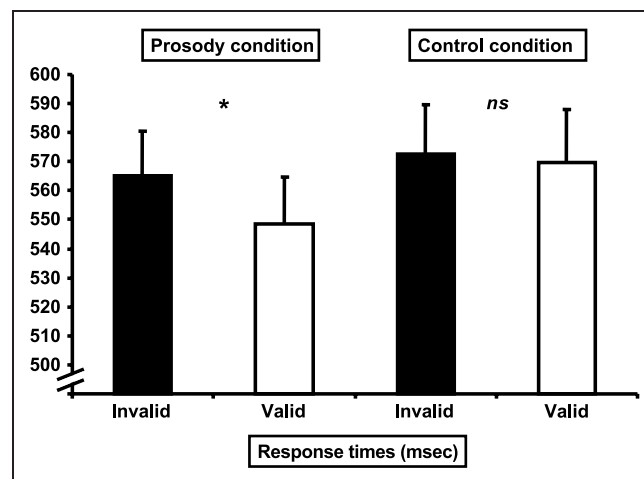


Figure 2. Response times (msec) for the prosody condition and the control condition. In the prosody condition, participants responded faster toward valid than invalid targets ($p = .06$). No such facilitation was observed for the control condition.

contrast to Brosch, Grandjean, et al. (2008), in the present study, participants responded only on 10% of the trials, as we wanted to analyze brain activity for the 90% of trials without contamination by motor responses. In the control condition, no differences were found in response times between valid (570 msec) and invalid (572 msec) trials [$t(14) = 0.4, ns$]. The interaction Voice condition \times Target position revealed longer response times toward targets presented to the left visual hemifield (580 msec) compared to the right visual field (562 msec) in the control condition [$F(1, 14) = 7.36, p = .02, \text{partial } \eta^2 = .35$].

ERP Analysis and Source Localization

Figure 3 shows the ERPs time-locked to target onset for targets presented to the left visual field and ERPs for the valid and invalid conditions for the prosody condition at electrodes PO7, PO8, O1, and O2.

P1 amplitude was larger in the prosody trials (3.0 μV) than in the control trials (1.9 μV), as revealed by the main effect of voice condition [$F(1, 14) = 68.98, p < .001$,

partial $\eta^2 = .83$]. P1 for targets presented to the right hemisphere peaked earlier (164 msec) than P1 for targets presented to the left hemisphere (171 msec), as indicated by a main effect of target position [$F(1, 14) = 15.14, p = .002, \text{partial } \eta^2 = .52$].

Most important for our hypotheses, the interaction Voice condition \times Cue validity was statistically significant [$F(1, 14) = 5.78, p = .03, \text{partial } \eta^2 = .29$]. We thus analyzed the data for the prosody condition and the control condition separately with regards to the effects of cue validity. In the prosody condition, amplitude of the P1 was larger in valid (3.2 μV) than in invalid (2.8 μV) trials as shown by a main effect of cue validity [$F(1, 14) = 6.82, p = .021, \text{partial } \eta^2 = .33$]. This effect was driven by targets presented to the left visual field (left visual field invalid: 2.6 μV , left visual field valid: 3.3 μV , right visual field invalid: 2.9 μV , right visual field valid: 3.0 μV), as indicated by the interaction Cue validity \times Target position [$F(1, 14) = 5.07, p = .041, \text{partial } \eta^2 = .27$] and a follow-up t test comparing valid and invalid targets presented to the left visual field [$t(14) = 3.9$,

Figure 3. Results from the ERP analysis: (A) ERPs time-locked to target onset for targets presented to the left visual field and ERPs for the valid (red) and invalid (black) conditions for the prosody condition at electrodes PO7, PO8, O1, and O2. (B) ERPs at O1 and O2 for the control condition.

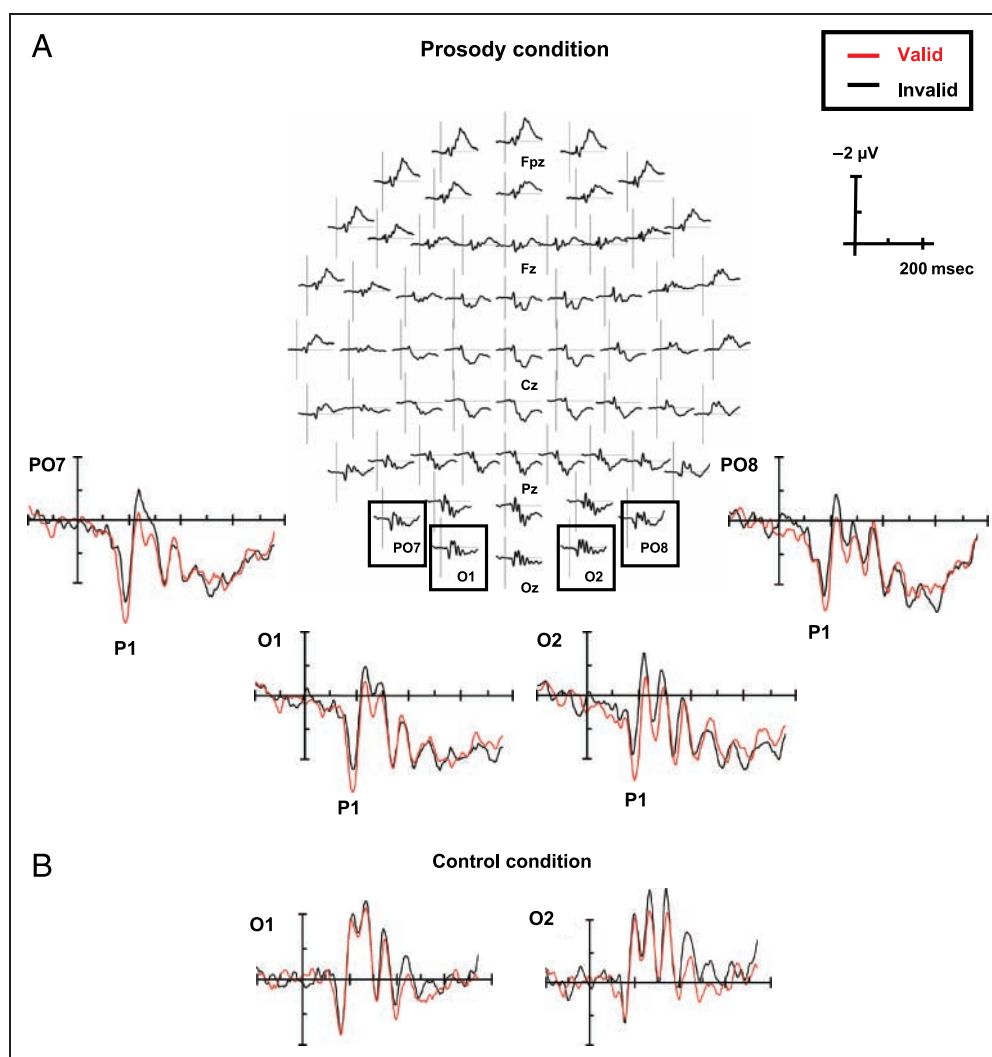
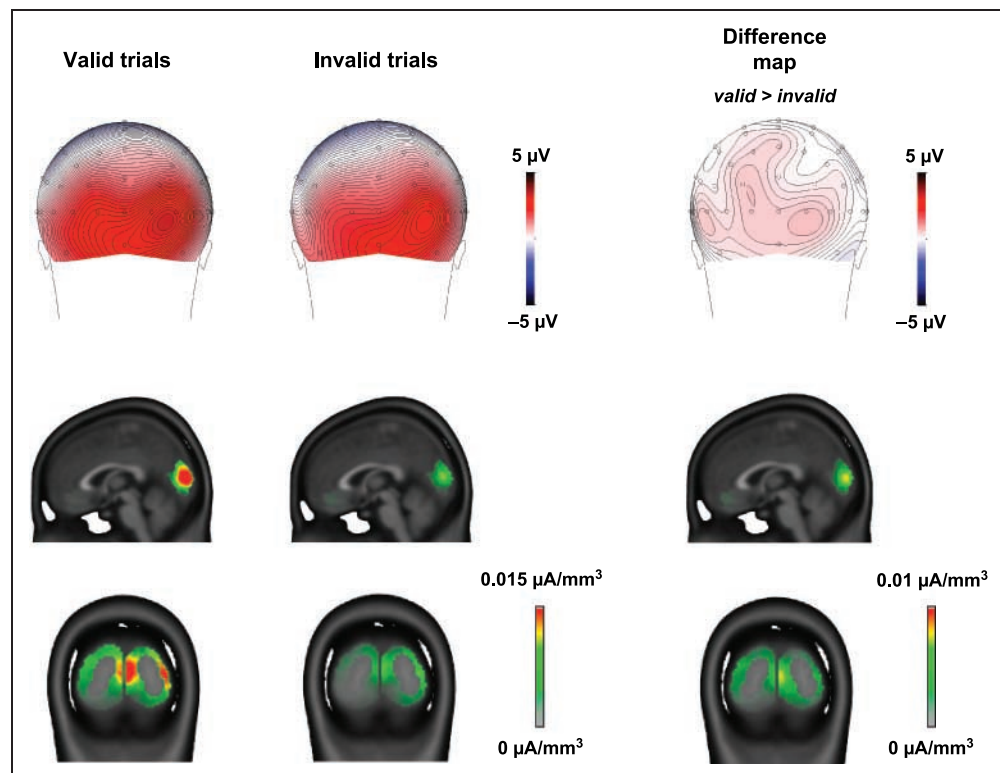


Figure 4. Top row: Topographic maps for the P1 in valid and invalid trials and topographic difference map. Middle and bottom rows: Inverse solution based on LAURA revealed the intracranial generators of the P1 in the striate and extrastriate visual cortex. Values of the inverse solution for valid and invalid trials are shown on a continuous scale from 0 to 0.015 $\mu\text{A}/\text{mm}^3$, for the difference map on a continuous scale from 0 to 0.01 $\mu\text{A}/\text{mm}^3$. A region-of-interest analysis taking into account the inverse solution points in the peak activation in the visual cortex confirmed stronger activation to valid than to invalid targets.



$p = .001$, one-tailed]. In the control condition, no effect involving cue validity was significant (all $p > .17$, left visual field invalid: 2.0 μV , left visual field valid: 2.1 μV , right visual field invalid: 1.8 μV , right visual field valid: 1.6 μV).

Finally, we applied an inverse solution on the basis of LAURA to the peak of the P1 potential for valid and invalid trials in the prosody condition. Results confirmed that the intracranial generators of the P1 were located in the striate and extrastriate visual cortex (see Figure 4), a pattern of brain regions which has been repeatedly found when looking at the generators of this early visual response (Noesselt et al., 2002; Martinez et al., 1999). A region-of-interest analysis, based on the inverse solution points in the peak activation in the visual cortex (see Figure 4), confirmed stronger activation to valid (0.015 μV) than to invalid (0.010 μV) targets [main effect cue validity: $F(1, 14) = 11.01$, $p = .005$, partial $\eta^2 = .44$].

DISCUSSION

During this cross-modal emotional dot-probe task, we recorded scalp ERPs to investigate at what stage of stimulus processing the deployment of visuospatial attention toward simple nonemotional visual targets was affected by spatially congruent or incongruent emotional information conveyed in affective prosody. At the behavioral level, participants were faster to respond to the orientation of a visual target when it appeared at the spatial location of a previously presented utterance with anger prosody compared to neutral prosody. This result is

consistent with our previous behavioral findings (Brosch, Grandjean, et al., 2008), even though the effect in the present study was only marginally significant ($p = .06$), probably due to the lower number of GO trials requiring a manual response. Importantly, this cross-modal emotional effect was not present when using synthesized control stimuli matched for the mean fundamental frequency and the amplitude envelope of each vocal stimulus used in the experiment, ruling out the possibility that these low-level acoustic parameters trigger the cross-modal emotional effect.

Analysis of scalp ERPs revealed a selective modulation of the P1 component toward visual targets preceded by spatially congruent auditory cues conveying emotional prosody, which was restricted to targets presented to the left visual hemifield. P1 amplitude was higher when the visual target appeared at the location of the source of the anger compared to neutral prosody. This modulation of the P1 as a function of the affective prosody was not observed in the control condition. Thus, this P1 effect consecutive to visual target processing most likely depends upon the activation of voice-specific processes (Grandjean et al., 2005; Belin et al., 2000) and cannot be explained by the processing of a particular range of frequency or a specific amplitude contour in the auditory stimuli.

Here we show that the cross-modal modulation of spatial attention triggered by emotional prosody affected early sensory stages of visual processing. The observed modulation by emotional prosody took place earlier than the modulation observed with nonemotional auditory

cross-modal cues (McDonald & Ward, 2000), which emerged as an attentional negativity between 200 and 400 msec. McDonald and Ward (2000) interpreted the absence of a P1 modulation as suggesting that the cross-modal effects of an auditory event on visual processes are located after the initial sensory processing of visual information. In contrast to their finding, our results show a modulation during initial stages of visual processing caused here by emotional auditory cues. Two methodological differences between the study by McDonald and Ward (2000) and our study should be discussed when comparing the results. The former study used a modified exogenous cueing paradigm, where only one auditory cue was presented, whereas in our study, we presented two cues simultaneously in a modified dot-probe paradigm. However, one would expect a more exhaustive processing of the cue stimulus when it is presented without direct competition for processing resources, not when it has to compete with other stimuli. Thus, it is unlikely that this accounts for the differences in early perceptual processing. A second methodological difference concerns the SOA between the cue and the target: Whereas McDonald and Ward (2000) used SOAs between 100 and 300 msec, we used SOAs between 550 and 750 msec. Our choice of SOAs was motivated by the fact that prosody is mainly due to temporal changes such as variations in stress and pitch (Ladd, 1996), and thus, needs some time to unfold.

Assuming that the different results are not due to methodological differences between the studies, they might reflect fundamental differences in the processing of emotional and nonemotional stimuli. A system that prioritizes orienting toward emotionally significant stimuli, operating across modalities, might produce a different pattern of modulation and integration than a system for the prioritization of perceptually salient stimuli. The perception and evaluation of emotional stimuli involves the activity of neural structures, especially the amygdala (Vuilleumier, 2005; Sander, Grafman, & Zalla, 2003), which are not involved in the cueing of attention toward merely perceptually salient stimuli (Desimone & Duncan, 1995). The amygdala plays a crucial role in highlighting relevant events by providing both direct and indirect top-down signals in sensory pathways which modulate the representation of emotional events (Vuilleumier, 2005). Affective prosody leads to increased activation of the amygdala and the superior temporal sulcus (Grandjean et al., 2005; Sander & Scheich, 2001). Functional connections between the amygdala and the visual cortex have been observed in animal tracer studies (Freese & Amaral, 2005) and in humans using diffusion tensor MRI (Catani, Jones, Donato, & Ffytche, 2003). Furthermore, increased activation of the visual cortex when listening to emotional prosody (Sander et al., 2005) or familiar voices (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005) probably reflects a functional coupling between auditory and visual cortices that can

facilitate the visual processing of targets (Vuilleumier, 2005).

The behavioral effect as well as the modulation of the P1 component observed in our study might reflect a boosting of perceptual representation of the visual stimulus in occipital brain areas, here triggered by a preceding affective voice. This conjecture is substantiated by our source localization results, which clearly indicate that the P1 modulation originated from generators localized in the visual cortex. Based on previous anatomical evidence, we suggest that this enhanced occipital activation for visual targets preceded by valid emotional voice cues is probably driven by feedback connections from the amygdala to the visual cortex, including the primary visual cortex (Freese & Amaral, 2005; Vuilleumier, 2005; Catani et al., 2003).

Emotional prosody is generally processed by both hemispheres (Schirmer & Kotz, 2006; Van Lancker & Sidtis, 1992). Some particularly relevant acoustical features related to emotional prosody, however, seem to involve the right hemisphere to a greater extent and induce more stimulus-related processing in this hemisphere (Ross & Monnot, 2008), as shown by neuroimaging results (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Sander & Scheich, 2001) and behavioral studies such as the dichotic listening task (Carmon & Nachshon, 1973; Haggard & Parkinson, 1971). This lateralization is in line with our findings, which indicated that the modulation effect was mainly driven by targets presented to the left visual field, which are primarily processed by the right hemisphere.

Further studies might investigate the effect of different types of prosody (such as happy, surprised, or disgusted) on attentional modulation. As no difference in strength of amygdala activation is observed when comparing positive and negative prosody (Sander & Scheich, 2001), one would expect that our findings are not restricted to anger prosody, but can be generalized to different kinds of emotional prosody. We recently presented evidence for a similar generalization for the visual modality in form of rapid attentional modulation toward several different kinds of emotionally relevant stimuli (Brosch, Sander, et al., 2008).

To sum up, in this study we explored the effects of cross-modal emotional attention. Both behavioral and electrophysiological data converge on the central finding that emotional attention can also operate across two different sensory modalities by boosting early sensory stages of processing.

Acknowledgments

We thank Gilles Pourtois for valuable comments on a previous draft of the article. This work was supported by the National Centre of Competence in Research (NCCR) Affective Sciences, financed by the Swiss National Science Foundation (no. 51NF40-104897), and hosted by the University of Geneva.

Reprint requests should be sent to Tobias Brosch, Swiss Centre for Affective Sciences, University of Geneva, 7, Rue des Batoirs, 1205 Geneva, Switzerland, or via e-mail: Tobias.Brosch@unige.ch.

REFERENCES

- Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General*, *134*, 258–281.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
- Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2008). Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody. *Cognition*, *106*, 1497–1503.
- Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting towards positive emotional stimuli. *Psychological Science*, *19*, 362–370.
- Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... Attention capture by infant faces. *Emotion*, *7*, 685–689.
- Brosch, T., & Sharma, D. (2005). The role of fear-relevant stimuli in visual search: A comparison of phylogenetic and ontogenetic stimuli. *Emotion*, *5*, 360–364.
- Carmon, A., & Nachshon, I. (1973). Ear asymmetry in perception of emotional non-verbal stimuli. *Acta Psychologica*, *37*, 351–357.
- Catani, M., Jones, D. K., Donato, R., & Ffytche, D. H. (2003). Occipito-temporal connections in the human brain. *Brain*, *126*, 2093–2107.
- Davidson, R. J., Maxwell, J. S., & Shackman, A. J. (2004). The privileged status of emotion in the brain. *Proceedings of the National Academy of Sciences, U.S.A.*, *101*, 11915–11916.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Driver, J., & Spence, C. (1998). Crossmodal attention. *Current Opinion in Neurobiology*, *8*, 245–253.
- Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: Evidence from event-related brain potential studies. *Neuroscience and Biobehavioral Reviews*, *25*, 497–511.
- Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., et al. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *NeuroReport*, *17*, 249–253.
- Fox, E., Russo, R., & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cognition and Emotion*, *16*, 355–379.
- Freese, J. L., & Amaral, D. G. (2005). The organization of projections from the amygdala to visual cortical areas TE and V1 in the macaque monkey. *Journal of Comparative Neurology*, *486*, 295–317.
- Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in Brain Research*, *156*, 235–247.
- Grandjean, D., Sander, D., Lucas, N., Scherer, K. R., & Vuilleumier, P. (2008). Effects of emotional prosody on auditory extinction for voices in patients with spatial neglect. *Neuropsychologia*, *46*, 487–496.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*, 145–146.
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.
- Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C. M., & Landis, T. (2001). Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. *Brain Topography*, *14*, 131–137.
- Haggard, M. P., & Parkinson, A. M. (1971). Stimulus and task factors as determinants of ear advantages. *Quarterly Journal of Experimental Psychology*, *23*, 168–177.
- Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *353*, 1257–1270.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.
- Lipp, O. V., & Derakshan, N. (2005). Attentional bias to pictures of fear-relevant animals in a dot probe task. *Emotion*, *5*, 365–369.
- Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*, 432–440.
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*, 15–20.
- Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, *9*, 296–305.
- Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*, 364–369.
- McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science*, *11*, 167–171.
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, *115*, 2195–2222.
- Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. *Cognition and Emotion*, *13*, 713–740.
- Morris, J. S., Friston, K. J., Buchel, C., Frith, C. D., Young, A. W., Calder, A. J., et al. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain*, *121*, 47–57.
- Noesselt, T., Hillyard, S. A., Woldorff, M. G., Schoenfeld, A., Hagner, T., Jancke, L., et al. (2002). Delayed striate cortical activation during spatial attention. *Neuron*, *35*, 575–587.
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*, 466–478.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex*, *14*, 619–633.
- Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2006). Neural systems for orienting

- attention to the location of threat signals: An event-related fMRI study. *Neuroimage*, *31*, 920–933.
- Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional–anatomic organization in right hemisphere. *Brain and Language*, *104*, 51–74.
- Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*, 303–316.
- Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage*, *28*, 848–858.
- Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Cognitive Brain Research*, *12*, 181–198.
- Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). Oxford: Oxford University Press.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*, 1–22.
- Spieler, L., Meuli, R., & Clarke, S. (2007). Extinction of auditory stimuli in hemineglect: Space versus ear. *Neuropsychologia*, *45*, 540–551.
- Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective–prosodic stimuli by left- and right-hemisphere-damaged subjects. *Journal of Speech and Hearing Research*, *35*, 963–970.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, *17*, 367–376.
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–594.
- Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, *30*, 829–841.
- Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*, 1271–1278.
- Vuilleumier, P., & Schwartz, S. (2001). Beware and be aware: Capture of spatial attention by fear-related stimuli in neglect. *NeuroReport*, *12*, 1119–1122.
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 495–501.