Available online at www.sciencedirect.com
ScienceDirect
Cognition 106 (2008) 1497–1503
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody ☆

Tobias Brosch [a,b,*], Didier Grandjean [a,b], David Sander [a,b], Klaus R. Scherer [a,b]

[a] *Department of Psychology, University of Geneva, 40, Bd. du Pont d'Arve, 1205 Geneva, Switzerland*
[b] *Swiss Centre for Affective Sciences, University of Geneva, 7, Rue des Battoirs, 1205 Geneva, Switzerland*

Received 26 October 2006; revised 23 May 2007; accepted 25 May 2007

Abstract

Emotionally relevant stimuli are prioritized in human information processing. It has repeatedly been shown that selective spatial attention is modulated by the emotional content of a stimulus. Until now, studies investigating this phenomenon have only examined *within-modality* effects, most frequently using pictures of emotional stimuli to modulate visual attention. In this study, we used simultaneously presented utterances with emotional and neutral prosody as cues for a visually presented target in a *cross-modal* dot probe task. Response times towards targets were faster when they appeared at the location of the source of the emotional prosody. Our results show for the first time a cross-modal attentional modulation of visual attention by auditory affective prosody.

Keywords: Attention; Emotion; Cross-modal; Prosody

☆ This research was supported by the National Centre of Competence in Research (NCCR) for Affective Sciences, financed by the Swiss National Science Foundation.
* Corresponding author. Tel.: +41 22 379 98 12; fax: +41 22 379 98 44.
E-mail addresses: Tobias.Brosch@pse.unige.ch (T. Brosch), Didier.Grandjean@pse.unige.ch (D. Grandjean), David.Sander@pse.unige.ch (D. Sander), Klaus.Scherer@pse.unige.ch (K.R. Scherer).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2007.05.011

1. Introduction

Emotionally relevant stimuli are prioritized in human information processing (Vuilleumier, 2005) as shown by results such as automatic activation of the amygdala by emotional stimuli (e.g., Whalen et al., 1998), faster detection times for emotional than neutral stimuli in search tasks (e.g., Brosch & Sharma, 2005), and modulation of selective attention towards emotional stimuli (e.g., Pourtois, Grandjean, Sander, & Vuilleumier, 2004). Until now, studies investigating emotional modulation of spatial attention have only 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. The signals entering these different channels might originate from a common source. To receive maximal benefit from multimodal input, the brain must coordinate and integrate those channels. Indeed, using a dot probe paradigm, it has already been shown that visual attention is modulated by a unilaterally presented, emotionally neutral acoustic stimulus (Driver & Spence, 1998; Spence & Driver, 1997). However, to our knowledge, the effect of emotion in such a multimodal paradigm was never investigated. If, for example, I hear that somebody is shouting angrily at me, while somebody else speaks in an emotionally neutral tone, it will be adaptive to guide visual attention towards the angry person's location to appraise the situation and its consequences in greater detail (Scherer, 2001). To test the hypothesis that visual attention is modulated by the emotional prosody of the human voice, we adapted the dot probe paradigm (MacLeod, Mathews, & Tata, 1986), a paradigm that has been extensively used to investigate within-modality attention bias by emotional material, to investigate cross-modal attentional bias of simultaneously presented utterances with emotional and neutral prosody on visual attention.

2. Experiment

2.1. Participants

Thirty-nine participants (15 males, mean age 27.0 years, all right-handed, with reported normal hearing and normal or corrected-to-normal vision) were recruited on the premises of the University of Geneva and received 10,– CHF for their participation. Data from one male participant had to be excluded due to failure to follow instructions.

2.2. 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 ms (50% male speakers, 50% anger prosody) were extracted from a database of pseudo-sentences that had been acquired and validated in earlier work (Banse

& Scherer, 1996). The anger trials were directly adopted from the database, the neutral trials were selected from the "boredom" and "interest" stimuli, selecting the most neutral, on the basis of a judgment study, asking participants ($N = 15$) to assess the "neutrality" and the "emotionality" of these excerpts. 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). Sixteen participants (undergraduate students, 14 females) 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: $F(1,19) = 459.46$, $p < .001$], 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"]: $F(1,19) = 52.94$, $p < 0.01$]. All stimuli were combined to 180 stereophonically presented paired utterances with angry–neutral or neutral–neutral prosody. 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. Overall mean energy for the angry–neutral pairs was 64.44 dB, for the neutral–neutral pairs 64.45 dB ($t(178) = -.025$, $p = .98$). 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 non-individualized 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, 2006). 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 Fig. 1a). The experiment was controlled by *e-prime*®. The auditory cues were presented using Sony MDR-CD470® headphones. The visual targets were presented on a Panasonic TH-42PW5® 42 in. plasma screen.

2.3. Procedure and data analysis

Fig. 1a shows the experimental procedure. After a random interval of 500–1000 ms, a central fixation cross was presented for 500 ms. Immediately afterwards, one of the audio pairs was presented to the participant. Five hundred ms after the onset of the acoustic stimulation, a dot appeared for 30 ms either on the left or right side of the screen, at a distance of 45 cm to the fixation cross. The participants were seated at 100 cm from the screen. Thus the angle between the target dot and the fixation cross was 24°, which is equivalent to the synthesized location of the audio stimulus pairs. Participants were instructed to indicate whether the dot appeared in the

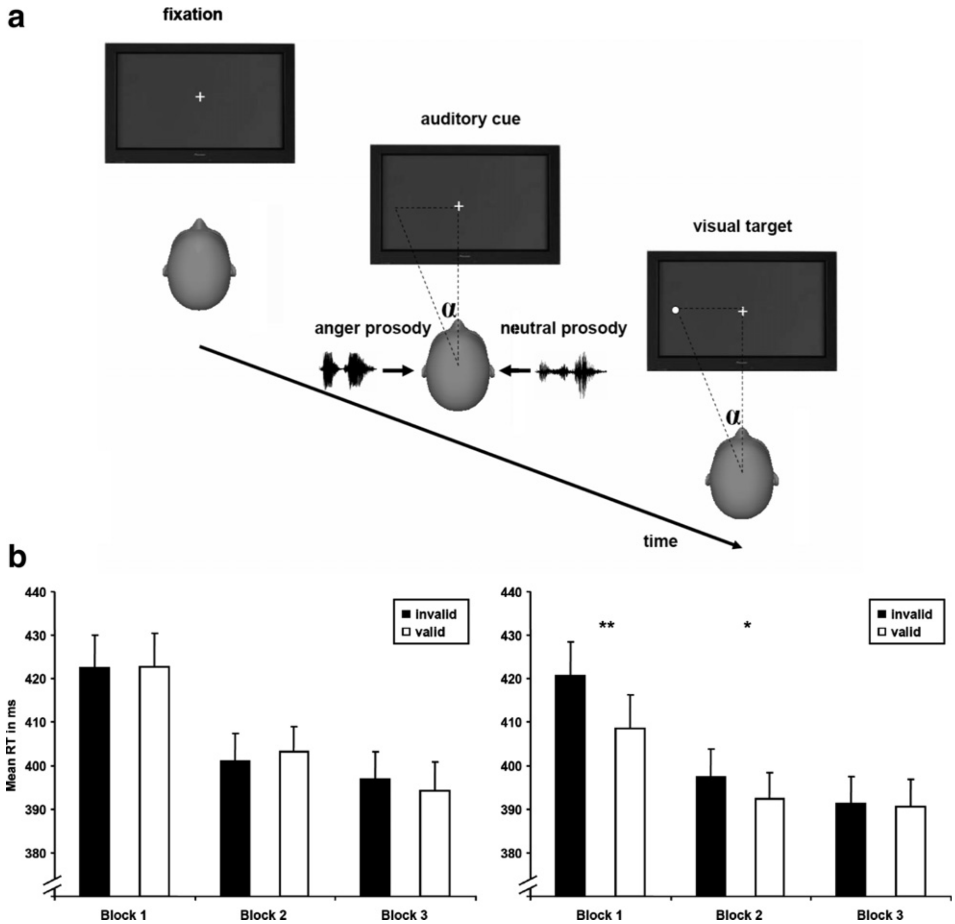

Fig. 1. (a) Experimental sequence for a valid trial, $\alpha = 24°$. (b) Response times in ms for invalid and valid trials presented to the left (left panel) and right (right panel) visual hemifield, $^{**}p = .002$, $^{*}p = .037$.

left or right part of the screen. Ten percent of the trials were trials without a dot to prevent participants from developing response strategies. Participants had a maximum of 1000 ms to respond, after that time, the next trial started. After a training phase of 8 trials, participants performed a total of 420 experimental trials. Only response times of correct responses lying within 3 SD around the mean for the respective individual were analyzed. Average error rate was 1.4%. Following the procedure in previous studies that used the dot probe task with visual emotional stimuli (Lipp & Derakshan, 2005), only the trials including angry–neutral cues were included in the analysis. In a *valid* trial, the dot appeared on the same side as the angry utterance, in an *invalid* trial, the dot appeared on the opposite side. Response times were analyzed in a $3 \times 2 \times 2$ – ANOVA design including the repeated factors *experimental block* (1st/2nd/3rd), *cue validity* (valid/invalid) and *visual hemifield of* *target* (left/right). When appropriate, Greenhouse-Geisser correction was used to correct the *F*-values in order to control for violations of sphericity.

2.4. Results

Fig. 1b shows the response times in milliseconds for targets presented to the left visual hemifield (left panel) and targets presented to the right visual hemifield (right panel). A main effect of *cue validity* indicated lower response times in valid compared to invalid trials, $F(1,37) = 8.82$, $p = .005$, partial $\eta^2 = .19$. A main effect of *experimental block* reflected faster responses during the course of the experiment, $F(2,74) = 30.29$, $p < .001$, $\varepsilon = .80$, partial $\eta^2 = .45$. This decrease in response time was linear across the blocks, as shown by the linear contrast, $F(1,37) = 38.77$, $p < .001$. A main effect of *visual hemifield of target* reflected faster responses to targets presented to the right visual hemifield compared to targets presented to the left visual hemifield, $F(1,37) = 4.23$, $p = .047$, partial $\eta^2 = .10$.

The interaction *cue validity* × *visual hemifield of target*, $F(1,37) = 5.27$, $p = .027$, partial $\eta^2 = .13$, indicated shorter RTs to targets presented to the right visual hemifield in valid trials (397 ms) compared to invalid trials (403 ms).[1] No cueing effects were apparent for targets presented to the left visual hemifield (valid: 407 ms, invalid: 407 ms). Furthermore, the three-way-interaction *experimental block* × *cue validity* × *visual hemifield of target* was statistically significant, $F(2,74) = 3.15$, $p = .049$, partial $\eta^2 = .08$. Post-hoc *t*-test showed that the cueing validity effect for targets presented to the right hemifield was significant during the first (invalid: 421 ms, valid: 409 ms, $t(37) = 3.21$, $p = .002$ one-tailed, partial $\eta^2 = .22$) and second block (invalid: 397 ms, valid: 392 ms, $t(37) = 1.85$, $p = .037$ one-tailed, partial $\eta^2 = .08$), but not during the third block of the experiment (invalid: 391 ms, valid: 390 ms, $t(37) = -.35$, n.s.). The decrease of the effect size for the cueing effect during the course of the experiment was linear, as indicated by the linear contrast, $F(1,37) = 7.15$, $p = .011$.

3. Discussion

In this experiment, we investigated whether emotional prosody influences the spatial deployment of visual attention in the dot probe paradigm when emotional and neutral utterances are presented simultaneously. Congruent with our hypothesis, the results show a cross-modal attentional modulation of visual attention by auditory affective prosody.

[1] When the neutral–neutral trials were included into the analysis, the interaction of *cue validity* × *visual hemifield of target* barely missed significance ($F(2,74) = 3.03$, $p = .054$). Post-hoc *t*-tests indicated significant differences between invalid (403 ms) and valid (397 ms) trials ($p = .002$) and between neutral (401 ms) and valid (397 ms) trials ($p = .018$) for right visual hemifield presentation. The three-way-interaction *experimental block* × *cue validity* × *visual hemifield of target* missed significance ($p = .16$).

The size of the attentional modulation effect decreased linearly during the experiment. This probably reflects habituation processes during the time course of the experimental sequence (see Breiter et al., 1996, for a comparable interpretation). Due to the repeated presentation of anger prosody without any threat-related consequences, the anger prosody might temporarily lose some of its signal value, which in turn might lead to a decreased allocation of attentional resources.

The effect was driven by targets presented to the right visual hemifield. A similar lateralization of the attentional modulation effect in the dot probe task has already been shown before by studies using the visual modality (Bradley, Mogg, Falls, & Hamilton, 1998; Bradley, Mogg, & Millar, 2000).

Our results mirror findings of visual attention modulation by visual emotional stimuli, but, for the first time, demonstrate this effect across modalities. It has been shown that the amygdala and the superior temporal sulcus are automatically activated by the anger prosody in a dichotic listening task (Grandjean et al., 2005; Sander et al., 2005). Our results might reflect this automatic activation on the behavioural level. Sander et al. (2005) observed an activation of the visual cortex when participants were attending to acoustic anger stimuli. They hypothesize that this might reflect a general boosting of sensory cortices as a response towards an attended emotionally relevant stimulus. Our result is in line with this speculation.

Further research should aim to investigate the effects of prosody expressing other emotions (e.g., joy or surprise) on visual spatial attention. In addition, the neural underpinnings and the exact time course of the observed attentional modulation should be explored.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.

Bradley, B. P., Mogg, K., Falls, S. J., & Hamilton, L. R. (1998). Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. *Cognition and Emotion, 12*, 737–753.

Bradley, B. P., Mogg, K., & Millar, N. H. (2000). Covert and overt orienting of attention to emotional faces in anxiety. *Cognition and Emotion, 14*, 789–808.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*, 875–887.

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.

Driver, J., & Spence, C. (1998). Crossmodal attention. *Current Opinion in Neurobiology, 8*, 245–253.

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.

Lipp, O. V., & Derakshan, N. (2005). Attentional bias to pictures of fear-relevant animals in a dot probe task. *Emotion, 5*, 365–369.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15–20.

Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex, 14*, 619–633.

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.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception and Psychophysics, 59*, 1–22.

Spierer, L., Meuli, R., & Clarke, S. (2006). Extinction of auditory stimuli in hemineglect: Space versus ear. *Neuropsychologia, 45*, 540–551.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9*, 585–594.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 18*, 411–418.