available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Human and animal sounds influence recognition of body language

Jan Van den Stock[a,b], *Julie Grèzes*[c], *Beatrice de Gelder*[a,d,*]

[a]*Cognitive and Affective Neuroscience Laboratory, Tilburg University, The Netherlands*
[b]*Old Age Psychiatry Department, University Hospitals Leuven, Belgium*
[c]*Laboratoire de Neurosciences Cognitives, INSERM U742 & DEC, Ecole Normale Supérieure, Paris, France*
[d]*Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts, USA*

ARTICLE INFO

Article history:
Accepted 15 May 2008
Available online 26 May 2008

Keywords:
Multisensory
Body perception
Emotion

ABSTRACT

In naturalistic settings emotional events have multiple correlates and are simultaneously perceived by several sensory systems. Recent studies have shown that recognition of facial expressions is biased towards the emotion expressed by a simultaneously presented emotional expression in the voice even if attention is directed to the face only. So far, no study examined whether this phenomenon also applies to whole body expressions, although there is no obvious reason why this crossmodal influence would be specific for faces. Here we investigated whether perception of emotions expressed in whole body movements is influenced by affective information provided by human and by animal vocalizations. Participants were instructed to attend to the action displayed by the body and to categorize the expressed emotion. The results indicate that recognition of body language is biased towards the emotion expressed by the simultaneously presented auditory information, whether it consist of human or of animal sounds. Our results show that a crossmodal influence from auditory to visual emotional information obtains for whole body video images with the facial expression blanked and includes human as well as animal sounds.

1. Introduction

When Hitchcock shows Norman Bates stabbing his victim to death in the shower or when the dorsal fin of a shark surfaces in "Jaws", the soundtrack is always there to underscore the message. Movie directors rely heavily on the extra dimension added to the movie experience by the soundtrack to convey emotion and aim at creating a multimodal experience in the viewer.

Experimental research on combined perception of auditory and visual stimuli has a long history (Müller, 1840), and there is now considerable evidence that multisensory stimuli presented in spatial or temporal proximity are bound by the brain into a unique perceptual gestalt (for reviews see de Gelder and Bertelson, 2003; Welch and Warren, 1986). Studies investigating the recognition of bimodal human emotional expressions typically consist of presenting audiovisual stimulus pairs in which the emotional content between the visual and auditory modality is either congruent or incongruent (de Gelder et al., 1999; de Gelder and Vroomen, 2000; Ethofer et al., 2006; Massaro and Egan, 1996; Spreckelmeyer et al., 2006; Van den Stock et al., 2007). For example, de Gelder and Vroomen

* *Corresponding author*. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, room 417, Building 36, First Street, Charlestown, Massachusetts 02129, USA. Fax: +1 31134662067.
E-mail address: degelder@nmr.mgh.harvard.edu (B. de Gelder).

Abbreviations: FMRI, functional magnetic resonance imaging; ERP, event-related potential

0006-8993/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2008.05.040

(2000) presented a static face expressing sadness or happiness combined with a spoken sentence with an emotionally neutral meaning but with either a sad or happy tone of voice. Participants were asked to ignore the voice and to indicate whether the face expressed happiness or sadness. The results indicated a clear crossmodal bias, e.g. a sad facial expression paired with a happy voice was recognized more as happy, compared to when the same facial expression was paired with a sad voice. In a follow up experiment, the task was reversed and participants were instructed to categorize the vocal expression and ignore the face. The results showed that the voice ratings were biased towards the emotion expressed by the face. The findings from de Gelder and Vroomen (2000) are consistent with other studies on bimodal perception of affect expressed in face and voice (de Gelder et al., 1999; Ethofer et al., 2006; Massaro and Egan, 1996).

We know from daily experience that emotions are not solely expressed in the face and the voice, but also conveyed very forcefully and over considerable distance by postures and movements of the whole body. Research on whole body perception is emerging as a new field in neuroscience (e.g. Atkinson et al., 2004; de Gelder, 2006; Grezes et al., 2007; Peelen and Downing, 2007). In view of these new findings a question is whether similar interactions as previously observed for facial expressions and auditory stimuli will also be obtained when observers are shown body–voice pairs. Recently, we presented static happy and fearful whole body expressions with faces blurred and each combined with a happy or fearful voice. Participants were asked to ignore the body expression and rate the emotion expressed by the voice. The results indicated that recognition of voice prosody was biased towards the emotion expressed by the whole body (Van den Stock et al., 2007, experiment 3). Here, we take that line of research a step further and investigate whether similar effects can be obtained with dynamic body images. Also, we address the question whether, as suggested by the familiar movie viewer's experience, there is crossmodal influence if both modalities are unmistakably and recognizably produced by a different source as is indeed often the case in naturalistic circumstances.

In this study, we present dynamic whole body expressions of emotion, showing persons engaged in an everyday activity and in a realistic context. In contrast to earlier studies we used non-verbal auditory information consisting of human vocalizations and also of animal sounds, two conditions that befit the naturalistic circumstances of viewing emotional body expressions from a relative distance. By using these two kinds of auditory information we address the issue whether environmental sounds (i.e. auditory stimuli originating from a source other than the visual stimulus) have a similar influence on recognition of visual human expressions as we expect voices to have.

Thirdly, to minimize semantic or verbal processing, which is initiated automatically when verbal information is presented, we used non-verbal auditory materials. Until now, only verbal vocalizations have been used to investigate crossmodal bias effects in processing human expressions. Non-verbal utterances have been used recently in scene–voice pairs. Spreckelmeyer et al. (2006) presented an emotionally sung syllable ("ha") paired with an emotional scene and asked participants to rate the valence of the scene. The authors did not observe an influence of the non-verbal vocalization on the ratings of the visual stimulus. However, pairing scenes with a sung syllable has limited ecological value. Also, a number of scenes in this study evoke an emotional experience, rather than showing an emotional expression (for example a picture of a baby or bunny).

Here, we investigate the influence of human and environmental emotional auditory information on the recognition of emotional body expression. For the case of the environmental auditory stimuli, we presented animal vocalizations inducing fear or happiness, creating realistic bimodal stimuli in the congruent conditions. Participants were presented video clips of happy or fearful body language. These were simultaneously presented with either congruent or incongruent human or animal vocalizations, or without auditory information. The experiment used a two alternative forced choice task and the instructions requested the participants to categorize the emotion expressed by the body stressing speed and accuracy.

2. Results

Trials with reaction times below 1000 ms and above 3000 ms (post-stimulus onset) were excluded. One participant responded outside this time window on more than 10% of the trials and was therefore excluded from the analysis. We computed the proportion happy responses of the different conditions. Results are shown in Fig. 1.

2.1. *Human vocalizations*

A repeated measures ANOVA was performed on the proportion happy responses with visual emotion (fearful and happy) and (human) auditory emotion (fearful, happy and no auditory stimulus) as within-subjects factors. This revealed a significant

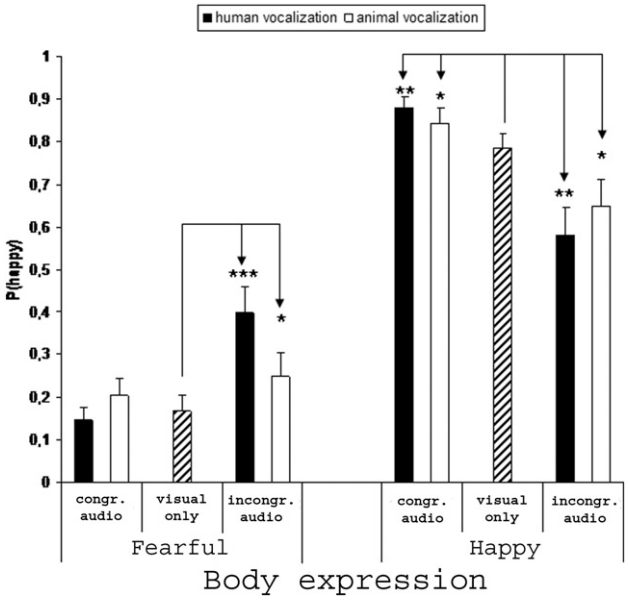

Fig. 1 – Proportion 'happy' responses in the bimodal and unimodal conditions, separated by emotion, auditory category and congruence. *$p<.05$; **$p<.01$; ***$p<.001$ congr.= congruent; incongr.= incongruent.