Brain Topogr (2009) 21:216–220
DOI 10.1007/s10548-009-0099-0

ORIGINAL PAPER

Instrumental Music Influences Recognition of Emotional Body Language

Jan Van den Stock · Isabelle Peretz · Julie Grèzes · Beatrice de Gelder

Received: 23 December 2008 / Accepted: 20 April 2009 / Published online: 7 May 2009

Abstract In everyday life, emotional events are perceived by multiple sensory systems. Research has shown that recognition of emotions in one modality is biased towards the emotion expressed in a simultaneously presented but task irrelevant modality. In the present study, we combine visual and auditory stimuli that convey similar affective meaning but have a low probability of co-occurrence in everyday life. Dynamic face-blurred whole body expressions of a person grasping an object while expressing happiness or sadness are presented in combination with fragments of happy or sad instrumental classical music. Participants were instructed to categorize the emotion expressed by the visual stimulus. The results show that recognition of body language is influenced by the auditory stimuli. These findings indicate that crossmodal influences as previously observed for audiovisual speech can also be obtained from the ignored auditory to the attended visual modality in audiovisual stimuli that consist of whole bodies and music.

Keywords Multisensory · Emotion · Body language · Music

This article is published as part of the Special Issue on Multisensory Integration.

J. Van den Stock · B. de Gelder
Laboratory of Cognitive and Affective Neuroscience, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

J. Van den Stock
Old Age Psychiatry Department, University Hospitals Leuven, Brusselsestraat 69, 3000 Leuven, Belgium

I. Peretz
Department of Psychology, BRAMS, University of Montreal, 1430 Boulevard du Mont-Royal, Outremont, Canada

J. Grèzes
Laboratoire de Neurosciences Cognitives, INSERM U742 & DEC, Ecole Normale Supérieure, Paris, France

B. de Gelder (✉)
Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, room 417, Building 36, First Street, Charlestown, MA 02129, USA
e-mail: degelder@nmr.mgh.harvard.edu

Introduction

The movie '2001: A Space Odyssey' is a landmark in the science-fiction genre. A classic scene shows a man-ape smashing a skeleton with a bone, while Richard Strauss's *Also Sprach Zarathustra* blasts in the background. It is the combination of these visual and auditory inputs that results in a unique experience in the viewer.

Research on multisensory perception has a long history (Müller 1840) and focussed on audiovisual speech (McGurk and MacDonald 1976). However, multisensory research on emotional events is scarce and was until recently limited to investigations into the perception of facial and vocal expressions (e.g., de Gelder and Vroomen 2000). In the latter type of studies two modalities are typically combined to create emotionally congruent and incongruent face–voice pairs and to provide a window into the integration process (de Gelder and Bertelson 2003). Participants are instructed to rate the emotion in one of the two modalities while ignoring the other. The results have shown that recognition of the emotion in the target modality is typically influenced towards the emotion expressed in the task irrelevant modality.

In two recent studies we have taken this issue beyond facial expressions and investigated affective crossmodal

influences in whole body expressions (Van den Stock et al. 2007, 2008). We investigated naturalistic actions that are part of everyday life and focussed on instrumental actions, like grasping and drinking. Our data showed that affective crossmodal effects occur with body–voice pairs, but also when body expressions are presented with animal vocalizations (Van den Stock et al. 2008). In the present study, we take the issue of affective crossmodal influence a step further and focus on bimodal stimuli that are not normally associated with each other, namely instrumental classical music and a person involved in an everyday action in a natural location.

Methods

Participants

Fifteen adult participants (7 male) were recruited through the local newspaper and were paid 20€. Mean age (SD) was 44.0 (12.6) years. Guidelines of the Declaration of Helsinki were followed.

Materials

Visual materials consisted of video recordings of 12 actors (6 male) who performed an everyday action (picking up a glass, drinking from it and putting it back on the table) and were shown in full body view. See Fig. 1 for examples. They performed this action with different emotional expressions (anger, disgust, fear, happiness, sadness and neutral). Before the performance, they were briefed with a specific scenario. For example, the happy scenario specified that the glass contained the favourite drink of the actor. The scenarios for all emotions are shown in Table 1. 3,000 ms fragments were taken from the recordings and the faces of the actors were blurred. We used video editing software (Adobe Aftereffects 8.0) to track the trajectory of the face in the movie and we replaced the face by a blurred mask. In a pilot study, all edited stimuli were presented four times in random order to 14 participants. They were instructed to categorize the emotion expressed by the actor in a six alternative forced choice task (anger, disgust, fear, happiness, sadness and neutral). On the basis of these, we selected 10 happy videos (5 male) and 10 sad videos (5 male) that were all correctly recognized above 75%.

Auditory materials consisted of fragments from the classical repertoire that expressed a happy or a sad tone and are described and validated in Peretz et al. (1998). We selected a 3,000 ms fragment from 10 happy and 10 sad excerpts. Results in Peretz et al. (1998) show that distinct happy and sad ratings are already elicited after 500 ms.

Procedure

Each of the 20 auditory stimuli was randomly paired once with a happy video and once with a sad video. This resulted in 40 unique bimodal stimuli of which 20 were congruent (e.g., happy audio paired with happy video) and 20 were incongruent (e.g., happy audio paired with sad video).

The experiment consisted of a visual (V), an auditory (A) and an audiovisual (AV) block. The order of blocks was randomized. A trial always started with presentation of a white fixation cross against a dark background shown for a variable duration of 1,000–3,000 ms to reduce temporal predictability. This was followed by presentation of a

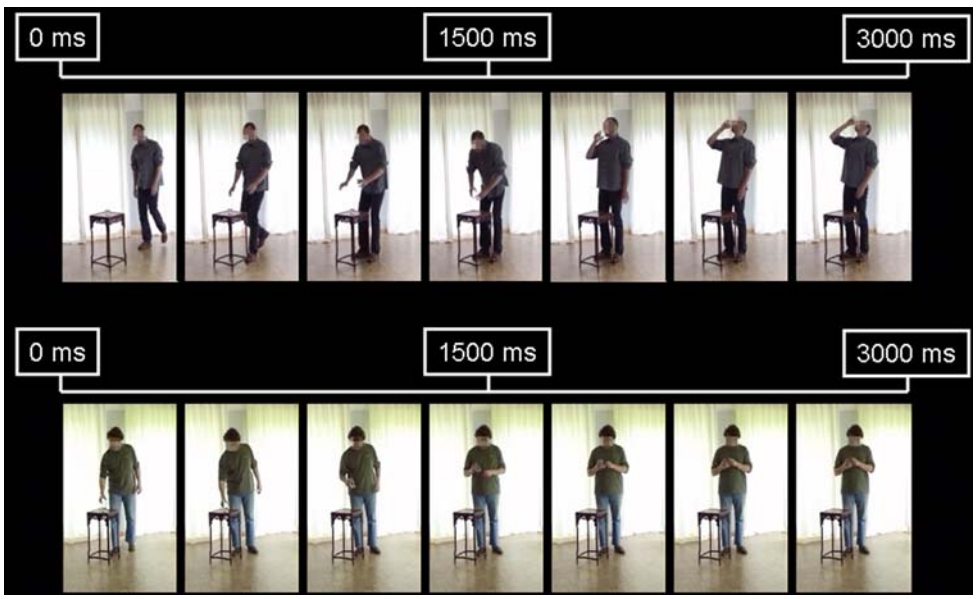

Fig. 1 Examples of frames from the video clips. The top row shows frames from a happy video, the bottom row shows frames from a sad video

Table 1 Scenarios provided to the actors to perform the action with different emotions

Emotion	Scenario
Anger	You just had an intense quarrel with a friend and you are very angry when you drink from the glass
Disgust	The glass contains a stomach-turning liquid and you are very disgusted
Fear	The content of the glass is extremely hot and you are afraid to drink it
Happiness	The glass contains your favourite drink and this makes you very happy
Sadness	You are returning from the funeral of a loved one and feel very sad, while you drink the glass
Neutral	Drink from the glass without any specific emotional state of mind

stimulus (V, A or AV 3,000 ms) after which a dark screen with a white question mark was shown until a response was given. In the V-block, the stimuli consisted of the 20 videos that were randomly presented one by one while the participants were instructed to categorize the emotion expressed by the actor. In the A-block the 20 auditory stimuli were randomly presented under the instruction to categorize the emotion expressed by the music. In the AV-block, all 40 bimodal stimuli were randomly presented and the instructions explicitly stated to categorize as accurately and as fast as possible the emotion expressed by the body language of the actor. The on- and offset of the visual and auditory stimuli in the AV-block were synchronized. Auditory stimuli were delivered through pc speakers located on the left and right of the screen. The volume was set at a comfortable listening level.

Results

The results are displayed in Fig. 2. We calculated the proportion happy responses for every condition in every block. A paired samples *t*-test showed there was no significant difference between the proportion happy responses in the happy audio condition and the proportion sad responses in the sad audio condition ($t(14) = .168$, $p < .869$), indicating both expressions are equally well recognized. A repeated measures ANOVA on the proportion happy responses with visual emotion (2 levels: happy and sad) and auditory emotion (3 levels: happy, sad and no audio) as within-subjects factors revealed a significant main effect of visual emotion ($F(1,14) = 506$, $p < .001$) and of auditory emotion ($F(2,28) = 4.734$, $p < .017$). The interaction was not significant ($F(2,28) = .278$, $p < .760$). The main effect of visual emotion indicates that the proportion happy responses is higher in the conditions with happy body language. This is to be expected since the task involved categorization of the emotion expressed by the body. Therefore, the main effect of visual emotion merely indicates that happy body expressions are more frequently rated as happy compared to sad body expressions. The main effect of auditory emotion indicates that the proportion happy responses differs according to auditory condition. Pairwise post-hoc comparisons (LSD corrected) on the main effect of auditory emotion showed that the proportion happy responses is significantly higher in the conditions with happy audio, compared to both sad audio ($p < .039$) and to no audio ($p < .005$). This means that both visual conditions, namely happy and sad dynamic whole body expressions are categorized more frequently as expressing happiness when they are presented simultaneously with happy music compared to when each of these visual conditions are presented with sad music or without auditory information.

The absence of a visual emotion × auditory emotion interaction effect indicates that the effect of auditory information is of equal magnitude in both visual conditions. The results also show that the proportion happy responses is lowest in the condition with sad audio, but the difference with the other audio conditions was not significant ($p < .086$).

Since the task stated to respond when the question mark appeared, no reaction time data were analysed.

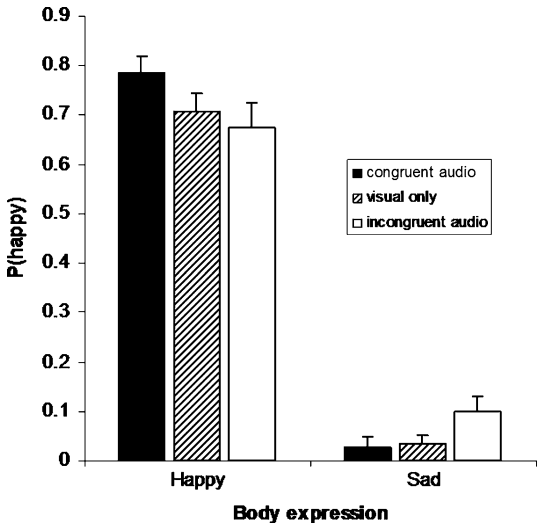

Fig. 2 Proportion 'happy' responses as a function of body expression and auditory information. Error bars represent 1 SEM

Discussion

The results show that emotional dynamic whole body expressions presented with happy music are recognized more as happy compared to when the same body expressions are paired with sad classical music or without auditory information. Even when instructions explicitly state to categorize the emotion expressed by the visual stimulus, the ratings are influenced towards the emotion expressed by the auditory stimulus. Our findings show that the influence of happy music is equally pronounced in both happy and sad body language. Moreover, body expressions presented with sad music are recognized more frequently as sad compared to when the bodies are presented in isolation or with happy music, but this effect is only marginally significant. The stronger influence of the auditory material on the happy body expressions compared to the sad expressions might be related to the level of intensity, valence and/or arousal expressed in the visual stimuli. Another possibility regards the matching of visual and auditory dynamics. It is not unlikely that congruence between changes in musical tempo and visual movement contributes to crossmodal influences. In previous studies, we have shown that whole body expressions of emotion can influence recognition of vocal emotional expressions (Van den Stock et al. 2007), but also that whole body expressions are influenced by both human and animal vocalizations (Van den Stock et al. 2008). The rate of co-occurrence of body–voice pairs in natural circumstances is high, since both are produced by the same source. Presumably the perceptual system is well versed in the simultaneous processing of jointly produced or at least naturalistically co-occurring visual and auditory inputs and may therefore rely on specialized mechanisms for cross-modal binding (de Gelder and Bertelson 2003). The combination of whole body expressions and animal vocalizations is less frequent and by this reasoning cross-modal influence between these stimulus categories is less evident. Still, simultaneous perception of a fearful body expression and a fear inducing dog bark can be perceived as one event, especially considered from an evolutionary perspective. However, the evolutionary significance of watching a person grasping an object while hearing instrumental classical music is less direct and less understood. The importance of the present findings lies in the fact that even multimodal inputs with no direct strong adaptive association can modulate the affective interpretation of clearly separate information streams. Nevertheless, instrumental music and body movements certainly occur frequently in dance, movies, social situations, etc. Our results show it is worth considering that the brain is organized for maintaining these flexible associations. Music may also mimic prosodic cues that otherwise communicate emotion vocally or through ambient environmental sounds. Even if the effectiveness of music for conveying emotion is entirely a learned process shaped by culture, it is interesting that the brain has found a way to link music to emotion and furthermore to cross-modally link music and bodily cues.

The pilot validation study consisted of a six alternative forced choice design. The primary aim of this pilot study was to assess how well the stimuli expressed the target emotion and therefore we offered the participants a range of response alternatives. In the main experiment, we choose to administer a design with two response alternatives in keeping with the design of our previous experiments (Van den Stock et al. 2007, 2008). The aim of the main experiment was to investigate crossmodal influence and we believe that a more limited number of response possibilities is preferable when making affective judgments in this context. Increasing the number of response alternatives may involve a higher appeal to more cognitive processes.

Despite the fact that the data from the pilot validation study show that the visual stimuli are easily recognizable when it comes to emotional categorization, one can not entirely exclude that the action in itself, i.e., drinking has an emotionally neutral association. For example it may be that drinking is associated with relief of thirst and is thereby biased towards a positive valence. However, the main interest of the present study concerns crossmodal influence elicited by the auditory information and this is measured by the difference between visual and audiovisual conditions. The primary focus of this study is the change between how congruent, incongruent and unimodal stimuli are categorized and the valence of the action itself (drinking) is equal in all the conditions.

One possible explanation for the observed effects might be that both visual and auditory emotional information elicit a similar affect program (Tomkins 1962, 1963; Panksepp 1998), which is neuro-anatomically supported by the involvement of premotor structures in perception of both body expressions (e.g. de Gelder et al. 2004) and music (e.g. Minati et al. 2008).

An alternative but not incompatible explanation at the neuro-anatomical level implies a link between production and perception of emotional actions. Bimodal mirror neurons in monkey premotor and motor structures display an increased firing rate when an action is either performed, seen or heard (Kohler et al. 2002). Indirect evidence from functional Magnetic Resonance Imaging (fMRI) supports the existence of a similar mirror (Grèzes et al. 2003) and bimodal mirror (Lahav et al. 2007) system in premotor cortex in humans. The latter study shows that the premotor cortex of non-musicians is more activated by listening to musical excerpts that they have recently learned to play on

the piano than by music they have never played. Although this study does not focus on the affective features of the music, it indicates that more complex auditory stimuli like classical music activate right premotor structures in humans. We have shown previously that perception of emotional body language also activates right premotor structures (de Gelder et al. 2004; Grezes et al. 2007). These combined findings may provide a neuro-anatomical framework to explain the crossmodal effects observed in the present study.

However, the focus here concerns the multimodal integration of *emotional information* and the role of the amygdala in processing emotional information has been well established (see Zald 2003 for a review). Moreover, previous studies using face–voice pairs have shown that crossmodal binding of affective information involves the left amygdala (Dolan et al. 2001; Ethofer et al. 2006) and this brain structure receives both visual and auditory inputs (McDonald 1998). Therefore, this may be a critical brain region involved in the unique experience one has when watching the ape-man in '2001: A Space Odyssey'.

The results of our behavioural study do not allow to formulate hypotheses about the perceptual underpinnings of the observed effects. We used a similar paradigm as our previous study (Van den Stock et al. 2008) but investigation of the stage at which affective crossmodal influence occurs, i.e., either a visual-perceptual level, a semantic post-perceptual level or even a response selection level, requires the complimentary use of imaging techniques, preferably with a high temporal resolution like electroencephalography (EEG) or magnetoencephalography (MEG).

Another issue concerns what it is in a dynamic whole body expression that makes it happy or sad. Sadness is typically more associated with lower muscle tonus and slower movements, whereas happiness usually involves quick and rapid movements, mostly involving raising of the arms. The movement parameters that are related to emotional communication have been extensively described earlier (e.g. Darwin 1872; Argyle 1988).

Our study makes a beginning with exploring how music influences the message conveyed by body language. The different levels at which music and body language make contact and the neurofunctional basis of our embodied music experience are just some of the many questions to be addressed in future research.

Acknowledgements We are grateful to A. Rous for help in data collection and to anonymous reviewers for helpful suggestions on the manuscript.

References

Argyle M (1988) Bodily communication. Methuen, London
Darwin C (1872) The expression of the emotions in man and animals. John Murray, London
de Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. Trends Cogn Sci 7:460–467
de Gelder B, Vroomen J (2000) The perception of emotions by ear and by eye. Cognition Emotion 14:289–311
de Gelder B, Snyder J, Greve D, Gerard G, Hadjikhani N (2004) Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. Proc Natl Acad Sci USA 101:16701–16706
Dolan RJ, Morris JS, de Gelder B (2001) Crossmodal binding of fear in voice and face. Proc Natl Acad Sci USA 98:10006–10010
Ethofer T, Anders S, Erb M, Droll C, Royen L, Saur R, Reiterer S, Grodd W, Wildgruber D (2006) Impact of voice on emotional judgment of faces: an event-related fMRI study. Hum Brain Mapp 27:707–714
Grezes J, Pichon S, de Gelder B (2007) Perceiving fear in dynamic body expressions. Neuroimage 35:959–967
Grèzes J, Armony JL, Rowe J, Passingham RE (2003) Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. Neuroimage 18:928–937
Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science 297:846–848
Lahav A, Saltzman E, Schlaug G (2007) Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci 27:308–314
McDonald AJ (1998) Cortical pathways to the mammalian amygdala. Prog Neurobiol 55:257–332
McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748
Minati L, Rosazza C, D'Incerti L, Pietrocini E, Valentini L, Scaioli V, Loveday C, Bruzzone MG (2008) FMRI/ERP of musical syntax: comparison of melodies and unstructured note sequences. Neuroreport 19:1381–1385
Müller JP (1840) Handbuch der Physiologie des Menschen. H. Ischer, Coblenz
Panksepp J (1998) Affective neuroscience: the foundation of human and animal emotions. Oxford University Press, New York
Peretz I, Gagnon L, Bouchard B (1998) Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. Cognition 68:111–141
Tomkins SS (1962) Affect, imagery and consciousness: the positive affects. Springer Verlag, New York
Tomkins SS (1963) Affect, imagery consciousness: Vol. 2. The negative affects. Springer verlag, New York
Van den Stock J, Righart R, de Gelder B (2007) Body expressions influence recognition of emotions in the face and voice. Emotion 7:487–494
Van den Stock J, Grezes J, de Gelder B (2008) Human and animal sounds influence recognition of body language. Brain Res 1242:185–190
Zald DH (2003) The human amygdala and the emotional evaluation of sensory stimuli. Brain Res Brain Res Rev 41:88–123