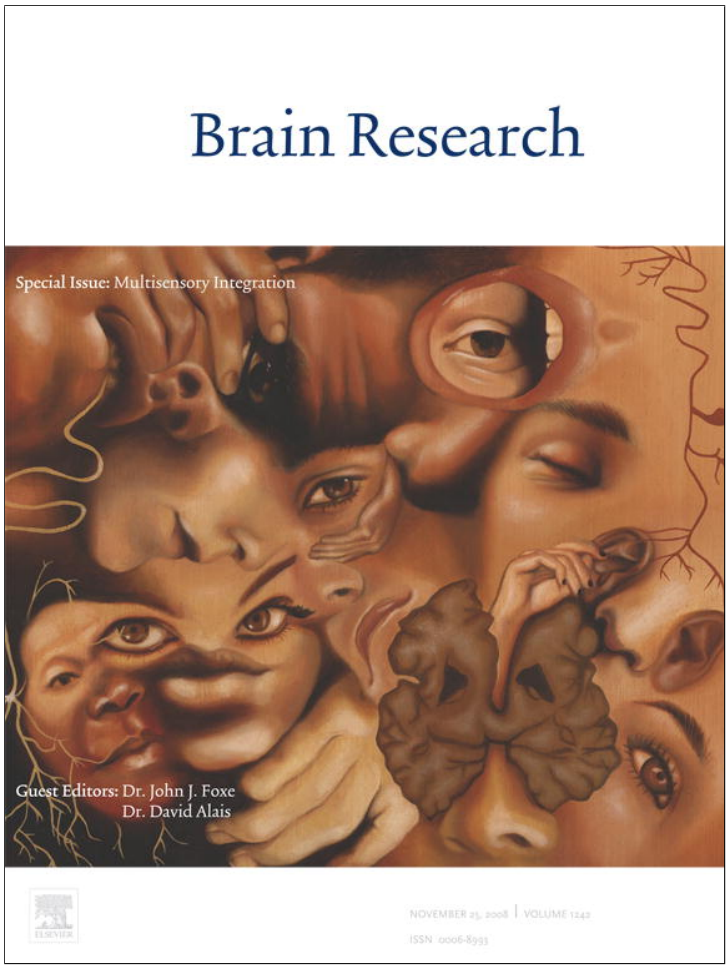

BRAIN RESEARCH 1242 (2008) 185–190

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

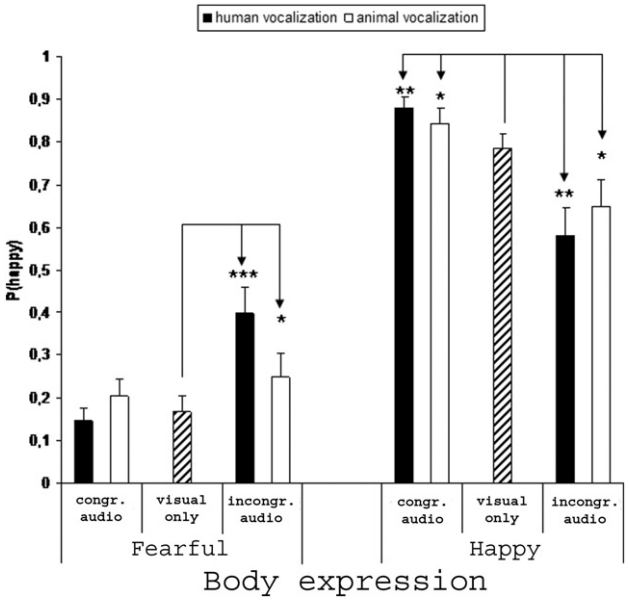

Fig. 1 – Proportion 'happy' responses in the bimodal and unimodal conditions, separated by emotion, auditory category and congruence. *$p<.05$; **$p<.01$; ***$p<.001$ congr.= congruent; incongr.= incongruent.

effect of visual emotion $F(1,25)=85.993$, $p<.001$, auditory emotion, $F(2,50)=16.523$, $p<.001$, and a significant interaction between visual emotion and auditory emotion, $F(2,50)=5.761$, $p<.006$.

To follow up on the interaction effect and to test the influence of the auditory stimuli on the recognition of the visual stimuli, we performed paired sample *t*-tests. Against the background of our previous experiments using faces and voices (e.g. de Gelder and Vroomen, 2000), we expect expression recognition performance on the congruent stimulus combinations to be better, compared to the unimodal combinations. Likewise, performance on the unimodal conditions is expected to be higher than on the incongruent conditions. Therefore, we performed one-tailed *t*-tests, comparing the unimodal conditions (V) with their respective bimodal (AV) conditions. For the happy body language, there was a difference between baseline [V(happy)] and both congruent AV, $t(25)=2.935$, $p<.01$, and incongruent AV, $t(25)=2.945$, $p<.01$. For the fearful body language, there was a significant difference between baseline [V(fear)] and incongruent AV, $t(25)=4.217$, $p<.001$.

2.2. *Animal vocalizations*

A repeated measures ANOVA on the proportion happy responses with visual emotion (fearful and happy) and (animal) auditory emotion (fearful, happy and no auditory stimulus) as within-subjects factors, revealed a significant effect of visual emotion $F(1,25)=92.050$, $p<.001$, auditory emotion, $F(2,50)=3.405$, $p<.041$, and an interaction between visual emotion and auditory emotion, $F(2,50)=5.040$, $p<.010$. The post-hoc paired *t*-tests (one-tailed) showed significant differences between [V (happy)] and congruent AV, $t(25)=1.823$, $p<.040$; between [V (happy)] and incongruent AV, $t(25)=1.948$, $p<.032$ and between [V(fear)] and incongruent AV, $t(25)=1.726$, $p<.050$.

2.3. *Human and animal vocalizations*

To compare the influence of human with animal vocalizations, we ran a 2 (video emotion: fearful and happy)×2 (auditory emotion: fearful and happy)×2 (auditory source: human and animal) repeated measures ANOVA on the proportion happy responses. This revealed a significant main effect of visual emotion $F(1,25)=56.048$, $p<.001$; auditory emotion $F(1,25)=11.001$, $p<.005$; a two-way visual×auditory emotion interaction $F(1,25)=11.564$, $p<.005$; a two-way auditory emotion×source interaction $F(1,25)=16.088$, $p<.001$; a two-way visual emotion×source interaction $F(1,25)=5.140$, $p<.05$; and a three-way visual emotion×auditory emotion ×source interaction $F(1,25)=5.532$, $p<.05$. The two-way auditory emotion ×source interaction indicates a different influence of the human and animal vocalizations. To follow up on this effect, we compared the influence of the human with the animal vocalizations, by computing the difference between the congruent and incongruent combinations, for the human and animal sounds separately (namely the human congruent conditions minus the human incongruent conditions and the animal congruent conditions minus the animal incongruent conditions). This difference was significantly larger for the human audio (mean 0.27, std 0.32) than for the animal audio (mean 0.12, std 0.31), as revealed by a two-tailed paired sample *t*-test $t(25) = 4.011$, $p<0.001$. The three-way interaction indicates the differential influence of the sources varies across visual emotion. We therefore computed the difference between the congruent and incongruent conditions for every auditory source and visual emotion. Paired *t*-tests showed for both happy and fearful body language a significant difference between the human congruent minus incongruent measure and the animal congruent minus incongruent measure.

Since a delayed reaction time task was used, no reaction time data were analyzed.

3. Discussion

The first aim of the present study was to investigate whether auditory information influences recognition of the emotion expressed in a simultaneously presented dynamic body expression. To test whether such crossmodal influence obtains, we presented video clips paired with non-verbal vocalizations and presented these stimuli with the instruction to categorize the emotion expressed by the body while ignoring the information provided by the auditory channel. Our results clearly indicate that recognition of body expressions is influenced by non-verbal vocal expressions. These findings are consistent with previous reports of crossmodal bias effects of vocal expressions on recognition of facial expressions, so far all using verbal stimuli (de Gelder et al., 1999; de Gelder and Vroomen, 2000; Ethofer et al., 2006; Massaro and Egan, 1996).

Our second aim was to investigate whether crossmodal influence is dependent on the perceived source of the auditory information or also obtains when different sources (human or animal sounds) have a similar signal function. Indeed, we find a clear influence of task irrelevant human voices on recognition of body language. However, the results also demonstrate that recognition of body language is influenced by environmental sounds. Happy body language is recognized better in combination with joyful bird songs, and recognized worse in combination with aggressive dog barks, compared to when the same happy body language is presented without auditory information. Human bodies are more intimately linked to human vocal expressions than animal vocalizations, which suggest that crossmodal influences are more probable in body–voice pairs, even if both can be perceived as carrying the same meaning, a typical example being danger signaling. The significant auditory emotion×source two-way interaction indicates that the impact of human vocalizations on the recognition of body language is larger than the impact of animal vocalizations. In view of the results of the pilot study which showed that human and animal vocalizations are recognized equally well, one may take this result as indicating that in general, human sounds influence recognition of human body language to a greater extent than animal sounds. Such an interpretation would be consistent with views in the literature on the importance of semantic and cognitive factors in multisensory pairing. A more finely tuned comparison of the impact of both sources would need a more detailed balancing of both sources, for example on the basis of the variability in pitch and volume. Much as such controls are needed in future research, we would like to point out that controlling the physical dimensions does not settle questions on the role of semantic and cognitive factors affecting crossmodal bias (de

Gelder and Bertelson, 2003). The nature of the privileged link between a facial or a bodily expression and auditory affective information produced by a person is at present not well understood. Similarly, comparisons between human sounds and the ones present in the environment have so far not been undertaken frequently. One recent suggestion is that the link between human face–body expressions and human vocalizations is based on premotor and motor structures in charge of producing the seen as well as the heard actions (Kohler et al., 2002). This would indeed explain the special status of human vocalizations observed here. But clear evidence in support of this view is currently not available. On the other hand, if at present there were a body of evidence, as for example could be provided by brain imaging studies, in support of the notion that heard and seen emotional expressions activate similar brain areas, alternative explanations come to mind. In fact, seen and heard emotion expressions may both activate the same "affect program" as argued for example by Tomkins (1962, 1963) and later Ekman (1982). Known convergence of auditory and visual afferents on the amygdala supports this view (Dolan et al., 2001). The latter alternative can accommodate easily the similarity in emotional signal function between human and animal sounds without appeal to a perception/production link. The present study raises these questions as topics for future research in the relatively novel field which will need to address the issues raised for three decades concerning the links between seen and heard speech perception. In the same vein future research will address the question whether the crossmodal bias also obtains between a visual image and a written word instead of its sound referent. This is again a matter that has been investigated in the area of audiovisual speech and been answered negatively (Vroomen and de Gelder, 2000).

The crossmodal influence we observe here is slightly different depending on whether the bodily expressions are fearful or happy. A comparison of the AV-conditions with the V-condition yielded a performance increase in AV-congruent condition and a performance decline in AV-incongruent condition for happy bodily expressions. For the fearful body language, we observe only a performance decline in AV-incongruent condition. So we find for the happy body language both a congruency and incongruency effect, but for the fearful body language, we find only an incongruence effect. The lack of a congruency effect for fearful body language cannot be explained by a ceiling effect given the results of the pilot data. We have currently no solid explanation for this differential crossmodal influence on the happy and fearful body language. An interesting topic for future research concerns the question whether the magnitude of emotional crossmodal influence differs between different emotions.

The results from the present study clearly indicate that crossmodal influences also occur even if both modalities are unmistakably produced by a different source. A relevant question would be what the conditions are for bimodal stimuli to be susceptible to crossmodal influences. Next to the obvious conditions of temporal and spatial congruence, animacy could play a role in the case of social stimuli. A recent event-related potential (ERP) study compared brain waveforms when perceiving human faces paired with either a human burp, or a monkey scream or a squeaking door. Results pointed to animacy specific neural responses, next to species-specific brain waveforms (Puce et al., 2007).

An important issue concerns the nature of the crossmodal influence. On the basis of a behavioral study, no direct inference can be made that the observed crossmodal influence has a perceptual basis. However, the instructions explicitly stated to base the emotional categorization solely on one modality (i.e. the visual), which is standard procedure in research dealing with conflicting multimodal inputs (Bertelson, 1998) and suggests an integrative perceptual process (de Gelder and Vroomen, 2000). Crossmodal integration of face–voice pairs seems unaffected by attentional resources (Vroomen et al., 2001) and the results of an ERP study indicate a very early integration of emotional faces and voices (around 110 ms after stimulus onset) (Pourtois et al., 2000). To examine the possible perceptual basis of a crossmodal bias effect with a behavioral paradigm, the ratings of unimodal stimuli in a pretest could be compared with the ratings of a post-test, with repeated presentations of bimodal pairs in between the pretest and post-test. The presence of after-effects of the bimodal presentations on the post-test unimodal ratings would point to a perceptual influence of the auditory information. The present study indicates the occurrence of crossmodal influences of both human and animal vocalizations on the recognition of dynamic body language, but does not allow conclusions concerning the nature of the effects.

Ecological validity is an important factor in multisensory integration (de Gelder and Bertelson, 2003). Multimodal inputs reduce stimulus ambiguity and the brain has primarily evolved to maximize adaptiveness in the real world, and this is one of the reasons why we choose visual stimuli with high ecological validity, namely the performance of an everyday action in the context of a realistic situation.

Recent functional magnetic resonance imaging (fMRI) studies looked at the neural correlates of integrating emotional faces and voices (Dolan et al., 2001; Ethofer et al., 2006; Kreifelts et al., 2007) and found increased activity in the left amygdala, when a fearful face was presented with a fearful voice (Dolan et al., 2001; Ethofer et al., 2006). The amygdala receive inputs from visual and auditory association cortex in the mammalian brain (McDonald, 1998) and its role in processing emotional stimuli is well established (see Zald, 2003 for a review). The amygdala therefore seems a primary candidate brain structure for integrating emotional information from different modalities.

4. Experimental procedures

4.1. *Participants*

Twenty-seven adults (14 male; 23 right-handed; mean age 31.5, range 18–50) participated in the experiment. They all gave written consent according to the Declaration of Helsinki. None of them had a history of neurological or psychiatric disorders. All had normal or corrected to normal vision and normal hearing.

4.2. *Stimulus materials*

4.2.1. *Visual stimuli*

Video recordings were made of 12 semi-professional actors (6 women), coached by a professional director. They were

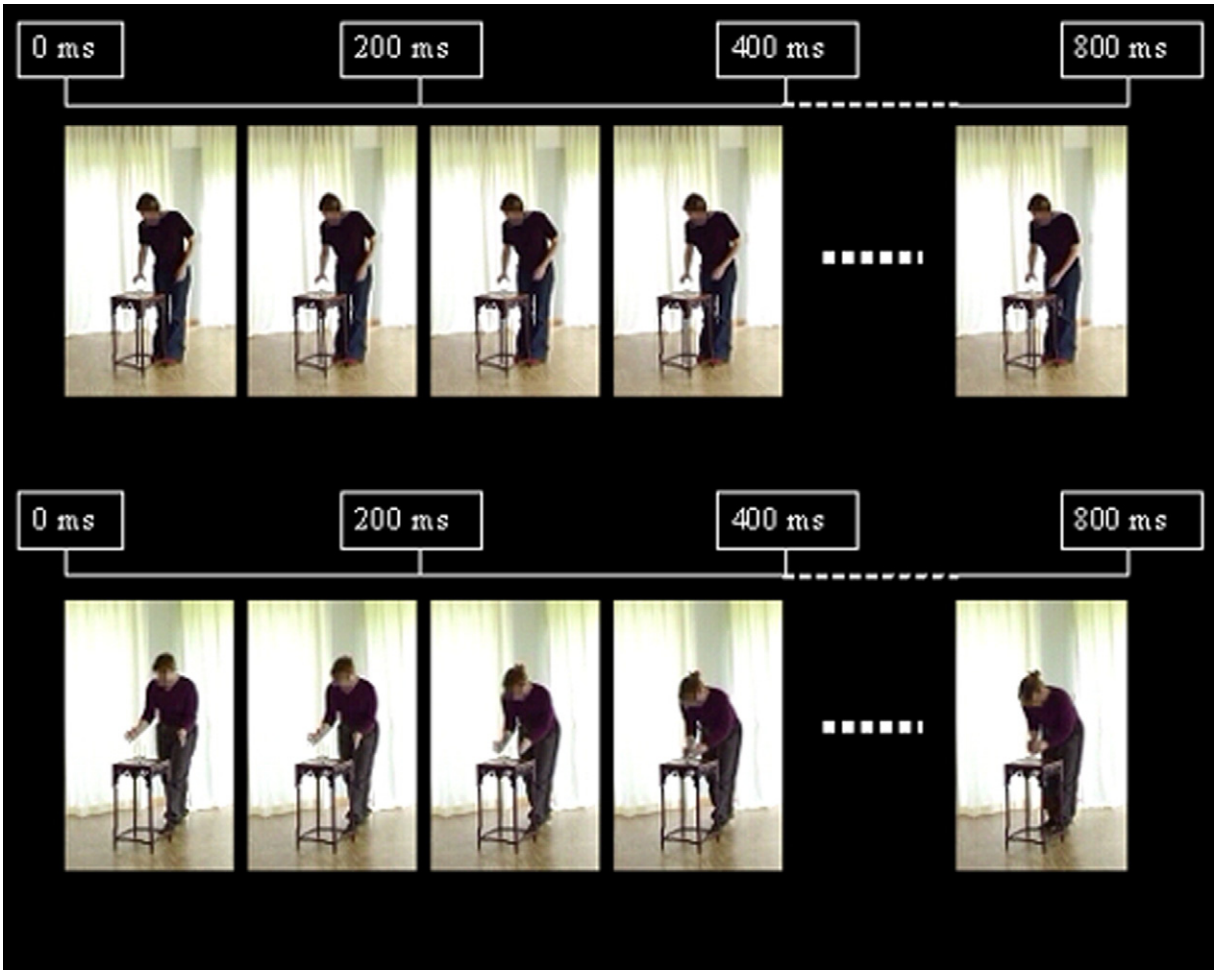

Fig. 2 – An example of frames from the video clips. The frame selection on the top row shows an actor grasping a glass in a fearful manner, the one on the bottom row performs the same action in a happy manner. The faces are blurred to minimize facial expression perception.

instructed to approach a table, pick up a glass, drink from it and to put it back on the table. They performed this action once in a happy and once in a fearful manner. A continuous fragment of 800 ms was selected from each video showing the actor grasping the glass. Facial expressions were blurred using motion tracking software. See Fig. 2 for an example.

In a pilot study the 24 edited dynamic stimuli (20 frames) were presented 4 times to 14 participants. Participants were instructed to categorize as accurately and as fast as possible the emotion expressed by the actor (fear or happiness). The pilot session was preceded by eight familiarization trials. Sixteen stimuli were selected (2 gender × 4 actors × 2 emotions). Since we expected that recognition of the body language improves when the body stimuli are combined with congruent auditory information, body stimuli that were recognized at ceiling were not selected. Mean recognition of the selected stimuli was 86.1% (SD 9.7). A paired *t*-test between the fearful and happy body language showed no significant difference, $t(13) = 1.109$, $p < .287$.

4.2.2. *Auditory stimuli*

Audio recordings were made at a sampling rate of 44.1 kHz of 22 subjects (14 women), while they made nonverbal emotional vocalizations (fearful and happy). Specific scripts were provided for every target emotion. For example, for fear the actors were instructed to imagine they were going to be attacked by a robber and to react to such an encounter in a non-verbal manner. Audio tracks were edited and the most representative 800 ms fragment from each recording was cut and digitally stored. In a pilot study the sounds were presented to 15 participants. Every sound was presented 4 times in a randomized order. The participants were instructed to categorize as accurately and as fast as possible the emotion expressed by the voice (fear or happiness). The pilot session was preceded by 3 familiarization trials. Based on these results, eight fearful and eight happy sounds were selected. Mean recognition of the stimuli was 94.6% (SD 6.7). A paired *t*-test between the fearful and happy vocalizations showed no significant difference, $t(14) = 0.474$, $p < .643$.

Environmental sounds consisted of aggressive dog barks and joyful bird songs and were downloaded from the internet. Stimuli were selected on the basis of their emotion inducing characteristics. In a third pilot study, these sounds were presented 4 times to 13 participants. They were instructed to categorize as accurately and as fast as possible the emotion induced by the sound (fear or happiness). The pilot session

was preceded by 3 familiarization trials. Eight fear inducing and eight happiness inducing sounds were selected. Mean recognition of the stimuli was 94.8% (SD 5.7). A paired *t*-test between the fearful and happy vocalizations showed no significant difference, $t(12) = 1.469$, $p < .168$.

For each emotion we compared the ratings of the animal vocalizations with those of the human vocalizations. Independent samples *t*-tests showed no differences between the pairs $t(26) \leq 1.195$, $p < .243$.

Experimental stimuli were then constructed with these visual and auditory materials. For this purpose each video file was paired once with a fearful and happy human vocalization, resulting in a total of 32 bimodal stimuli (human video/human audio) and once with a fear (dog barking) and happiness (birdsong) inducing animal vocalization, resulting in a total of 32 bimodal stimuli (human video/animal audio).

4.3. *Procedure*

The experiment consisted of a visual (V) and an audio-visual (AV) block. In each block all stimuli were presented twice in random order. The order of the blocks was counterbalanced. The AV-block consisted of 128 trials (2 presentations of 64 stimuli: 16 fearful videos with congruent human sounds, 16 fearful videos with incongruent human sounds, 16 videos with congruent animal sounds and 16 videos with incongruent animal sounds), the V-block of 32 trials (2 presentations of 16 stimuli, 8 fearful and 8 happy clips). A trial started with the presentation of a white fixation cross in the center of the screen against a dark background. The fixation cross had a variable duration to reduce temporal predictability (2000–3000 ms) and was followed by presentation of a stimulus (800 ms) after which a question mark appeared until the participant responded. A two alternative forced choice task was used requiring the participants to categorize the emotion expressed in the body by pressing the corresponding button (happy or fearful). Response buttons were counterbalanced across participants. Because we wanted to make sure participants saw the full length of the stimulus before they responded, they were instructed only to respond when the question mark appeared.

Acknowledgments

The research was partly funded by a Human Frontiers Science Program grant (RGP0054/2004-C), a European Commission grant (FP6-2005-NEST-Path Imp 043403-COBOL) and a NWO grant to BdG (Dutch Science Foundation).

REFERENCES

Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W., 2004. Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33 (6), 717–746.

Bertelson, P., 1998. Starting from the ventriloquist: the perception of multimodal events. In: Sabourin, M., Craik, I., Roberts, M. (Eds.), Advances in Psychological Science. Biological and Cognitive Aspects, vol. 2. Lawrence Erlbaum, Hove, UK, pp. 419–439.

de Gelder, B., 2006. Towards the neurobiology of emotional body language. Nat Rev., Neurosci. 7 (3), 242–249.

de Gelder, B., Bertelson, P., 2003. Multisensory integration, perception and ecological validity. Trends Cogn. Sci. 7 (10), 460–467.

de Gelder, B., Bocker, K.B., Tuomainen, J., Hensen, M., Vroomen, J., 1999. The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. Neurosci. Lett. 260 (2), 133–136.

de Gelder, B., Vroomen, J., 2000. The perception of emotions by ear and by eye. Cogn. Emot. 14 (3), 289–311.

Dolan, R.J., Morris, J.S., de Gelder, B., 2001. Crossmodal binding of fear in voice and face. Proc. Natl. Acad. Sci. U. S. A. 98 (17), 10006–10010.

Ekman, P., 1982. Emotion in the Human Face. Cambridge University Press, Cambridge.

Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., et al., 2006. Impact of voice on emotional judgment of faces: an event-related fMRI study. Hum. Brain Mapp 27 (9), 707–714.

Grezes, J., Pichon, S., de Gelder, B., 2007. Perceiving fear in dynamic body expressions. NeuroImage 35 (2), 959–967.

Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297 (5582), 846–848.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D., 2007. Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. NeuroImage 37 (4), 1445–1456.

Massaro, D.W., Egan, P.B., 1996. Perceiving affect from the voice and the face. Psychon. Bull. Rev. 3, 215–221.

McDonald, A.J., 1998. Cortical pathways to the mammalian amygdala. Prog. Neurobiol. 55 (3), 257–332.

Müller, J.P., 1840. Handbuch der physiologie des menschen. Coblenz: H. Ischer.

Peelen, M.V., Downing, P.E., 2007. The neural basis of visual body perception. Nat. Rev., Neurosci. 8 (8), 636–648.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M., 2000. The time-course of intermodal binding between seeing and hearing affective information. NeuroReport 11 (6), 1329–1333.

Puce, A., Epling, J.A., Thompson, J.C., Carrick, O.K., 2007. Neural responses elicited to face motion and vocalization pairings. Neuropsychologia 45 (1), 93–106.

Spreckelmeyer, K.N., Kutas, M., Urbach, T.P., Altenmuller, E., Munte, T.F., 2006. Combined perception of emotion in pictures and musical sounds. Brain Res. 1070 (1), 160–170.

Tomkins, S.S., 1962. Affect, Imagery and Consciousness: The Positive Affects. Springer Verlag, New York.

Tomkins, S.S., 1963. Affect, Imagery Consciousness. The Negative Affects, vol. 2. 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 (3), 487–494.

Vroomen, J., de Gelder, B., 2000. Crossmodal integration: a good fit is no criterion. Trends Cogn. Sci. 4 (2), 37–38.

Vroomen, J., Driver, J., de Gelder, B., 2001. Is cross-modal integration of emotional expressions independent of attentional resources? Cogn. Affect Behav. Neurosci. 1 (4), 382–387.

Welch, R.B., Warren, D.H., 1986. Intersensory interactions. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Performance. Sensory Processes and Perception, vol. 1. John Wiley and Sons, New York, pp. 25.21–25.36.

Zald, D.H., 2003. The human amygdala and the emotional evaluation of sensory stimuli. Brain Res. Brain Res. Rev. 41 (1), 88–123.