

# Impact of Voice on Emotional Judgment of Faces: An Event-Related fMRI Study

Thomas Ethofer,<sup>1,2\*</sup> Silke Anders,<sup>1,2</sup> Michael Erb,<sup>1</sup> Christina Droll,<sup>2</sup>  
Lydia Royen,<sup>2</sup> Ralf Saur,<sup>1</sup> Susanne Reiterer,<sup>1,2</sup> Wolfgang Grodd,<sup>1</sup>  
and Dirk Wildgruber<sup>2</sup>

<sup>1</sup>Section of Experimental MR of the CNS, Department of Neuroradiology, University of Tuebingen, Tuebingen, Germany

<sup>2</sup>Department of General Neurology, Hertie-Institute for Clinical Brain Research, University of Tuebingen, Tuebingen, Germany

---

**Abstract:** Emotional information can be conveyed by various means of communication, such as propositional content, speech intonation, facial expression, and gestures. Prior studies have demonstrated that inputs from one modality can alter perception in another modality. To evaluate the impact of emotional intonation on ratings of emotional faces, a behavioral study first was carried out. Second, functional magnetic resonance (fMRI) was used to identify brain regions that mediate crossmodal effects of emotional prosody on judgments of facial expressions. In the behavioral study, subjects rated fearful and neutral facial expressions as being more fearful when accompanied by a fearful voice as compared to the same facial expressions without concomitant auditory stimulus, whereas no such influence on rating of faces was found for happy voices. In the fMRI experiment, this shift in rating of facial expressions in presence of a fearfully spoken sentence was correlated with the hemodynamic response in the left amygdala extending into the periamygdaloid cortex, which suggests that crossmodal effects on cognitive judgments of emotional information are mediated via these neuronal structures. Furthermore, significantly stronger activations were found in the mid-portion of the right fusiform gyrus during judgment of facial expressions in presence of fearful as compared to happy intonations, indicating that enhanced processing of faces within this region can be induced by the presence of threat-related information perceived via the auditory modality. Presumably, these increased extrastriate activations correspond to enhanced alertness, whereas responses within the left amygdala modulate cognitive evaluation of emotional facial expressions. *Hum Brain Mapp* 27:707–714, 2006. © 2006 Wiley-Liss, Inc.

**Key words:** affective prosody; amygdala; emotion; facial expression; fear; fusiform gyrus

---

## INTRODUCTION

Emotional information can be transmitted by verbal utterances or by nonverbal means of communication, such as facial expression, emotional prosody, and gestures. Studies on multimodal integration of emotional information revealed biases between facial expression and affective prosody that occur even under the explicit instruction to ignore one of the modalities [de Gelder and Vroomen, 2000; Massaro and Egan, 1996]. It has been shown that such crossmodal biases are not dependent on attentional resources, arguing for the mandatory nature of such effects [Vroomen et al., 2001.] So far, experimental designs of studies on audiovisual integration of emotional information have relied on two-alternative choice tasks [de Gelder and Vroomen, 2000; Massaro and Egan, 1996].

---

Contract grant sponsor: Deutsche Forschungsgemeinschaft; Contract grant number: WI 2101/1-1; Contract grant sponsor: Heidelberg Academy of Sciences and Humanities.

\*Correspondence to: Thomas Ethofer, MD, Section of Experimental MR of the CNS, Department of Neuroradiology, Otfried-Mueller-Strasse 51, University of Tuebingen, 72076 Tuebingen, Germany.

E-mail: Thomas.Ethofer@med.uni-tuebingen.de

Received for publication 23 March 2005; Accepted 12 August 2005

DOI: 10.1002/hbm.20212

Published online 12 January 2006 in Wiley InterScience (www.interscience.wiley.com).

It has been shown, however, that judgments of emotional stimuli are not carried out in a dichotomous way, but continuously along the two dimensions of valence and arousal [Bradley and Lang, 1994]. Moreover, psychophysiological reactions such as blink reflex and skin conductance response to such stimuli covary with evaluative judgments along these dimensions, underlining their biological relevance [Lang, 1995]. Categorization in one of two possible emotions therefore might not reflect common demands on emotional information processing in everyday life, where more fine-grained shifts in interpretation of nonverbal affective information may occur. To determine whether such shifts in interpretation of visually presented affective information can be introduced by nonverbal emotional information perceived via the auditory modality, we conducted a behavioral experiment in which subjects were to rate the valence of facial expressions in the presence and absence of an affectively spoken sentence on a nine-point self-assessment manikin (SAM) scale [Bradley and Lang, 1994].

Subsequently, the same experimental paradigm was used in a functional magnetic resonance imaging (fMRI) study to determine which brain structures mediate biases between auditory and visual nonverbal information. Previous neuroimaging studies demonstrated enhanced responses within the fusiform gyrus to emotional as compared to neutral facial expressions [Morris et al., 1998; Vuilleumier et al., 2001], which are especially pronounced if the displayed emotion signals fear [Surguladze et al., 2003]. These augmented responses in extrastriate visual cortex to signals of threat have been suggested to reflect the preferential direction of visual attention to imminent signals of danger [Surguladze et al., 2003]. Using fMRI, we wanted to clarify if such an increase of the hemodynamic response in extrastriate brain regions during visual processing can also be induced across modalities by a threat-related auditory stimulus.

It has been suggested that modulation of extrastriate cortical responses are mediated by the amygdala [Davis and Whalen, 2001; Freese and Amaral, 2005; Morris et al., 1998; Vuilleumier et al., 2004] via backward projections from the amygdala to the visual cortex [Pitkänen, 2000]. Furthermore, the amygdala is a candidate structure to mediate crossmodal effects of affective prosody on judgment of emotional facial expressions, because its basolateral nuclei receive highly processed inputs from both auditory and visual association cortices [McDonald, 1998] and might therefore serve as a link for convergence of audiovisual inputs of emotional significance. In line with this assumption, a recent fMRI study revealed enhanced activation of the amygdala to fearful faces when there was emotional congruency with the affective intonation of a concomitant voice [Dolan et al., 2001].

The present study was designed specifically to determine whether the amygdala mediates mandatory effects of an implicitly processed affective voice on explicit judgment of the emotional valence of facial expressions. We thus correlated the blood oxygenation level-dependent (BOLD) response induced by the presence of an emotional intonation

with its effect on valence rating of facial expressions. We also predicted that the hemodynamic response in the amygdala is correlated with the behavioral effect of an emotionally spoken sentence on the explicit judgment of facial expressions.

## MATERIALS AND METHODS

### Subjects

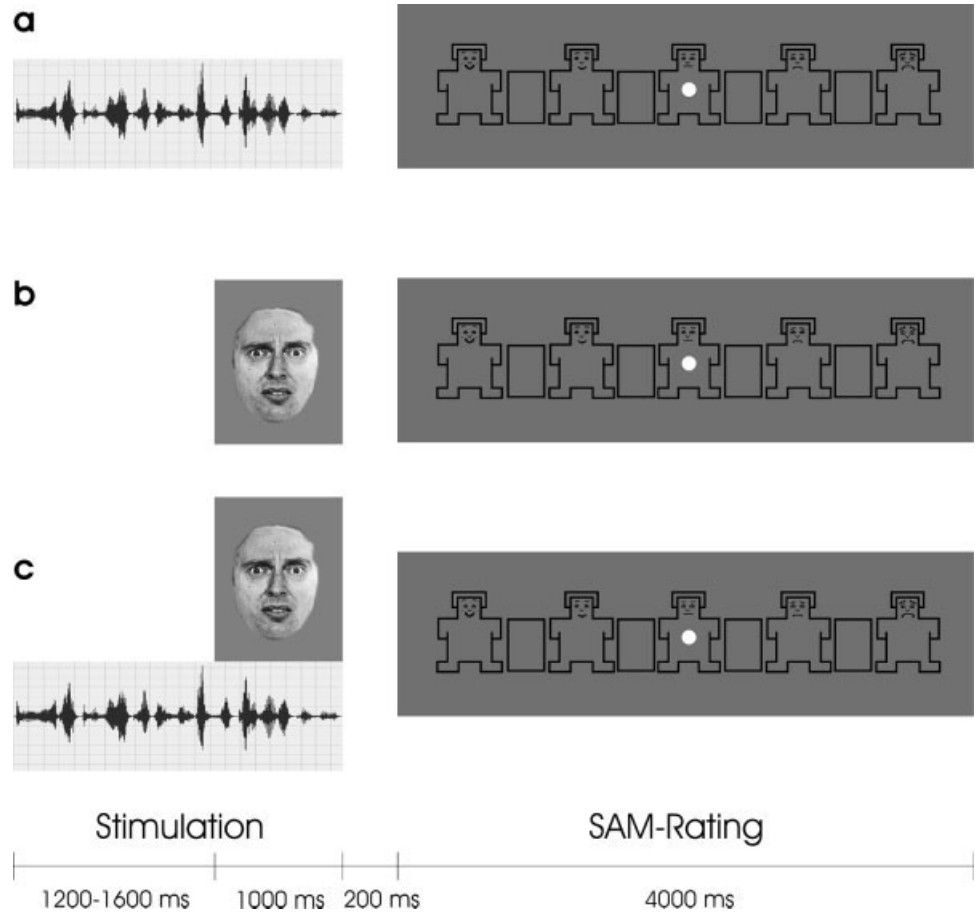
Thirty-three right-handed volunteers (16 men, 17 women; aged 18–35 years) participated in the behavioral experiment. Twelve right-handed subjects (seven men and five women; aged 19–29 years) who did not take part in the behavioral experiment were included in the fMRI experiment. None of the participants had a history of neurological or psychiatric illness. The study was compliant with the Code of Ethics of the World Medical association (Declaration of Helsinki) and the protocol of human investigation was approved by the local ethics committee. All subjects gave their written informed consent for the study.

### Stimuli

Visual stimuli consisted of 36 facial expressions of two male and two female Caucasian individuals (obtained from the Ekman and Friesen [1976] series) ranging from neutral to 100% happiness and neutral to 100% fear in incremental steps of 25% created by digital morphing techniques [Perrett et al., 1994]. Auditory stimuli included German sentences consisting of 14 syllables each (duration of 2.2–2.7 sec), which were emotionally neutral with respect to their semantic content (e.g., “Der Gast hat sich für Donnerstag ein Zimmer reserviert,” “The guest reserved a room for Thursday”). Two professional actors (one male and one female) pronounced 20 sentences in happy, fearful, sad, disgusted, and angry intonation. To select stimuli that most clearly represent the intended emotional prosody, the complete set of stimuli (20 sentences  $\times$  five emotions  $\times$  two actors) was presented in a pretest to 10 healthy subjects (five females, five male; aged 22–28 years) together with a tape recording of scanner noise to simulate the acoustic environment of an fMRI experiment [see Wildgruber et al., 2005]. From this stimulus set, 18 sentences spoken in happy and fearful prosody that were identified correctly by at least 70% of the subjects in the pretest were selected for the behavioral and the fMRI experiment.

### Experimental Design

Both the behavioral and the fMRI experiment comprised six sessions, each of which included 36 trials in a pseudorandom order. In two of these sessions, every trial consisted of a presentation of a facial expression shown for 1 sec without any auditory accompaniment, in another two sessions only auditory stimuli were presented, and in the remaining two sessions auditory and visual stimuli were presented with facial expressions being shown during the last second of the spoken sentences. In the unimodal visual and



**Figure 1.**

Experimental setup comprising two auditory (a), two visual (b), and two bimodal (c) sessions. In all three conditions, 200 msec after stimulus offset a nine-point self-assessment manikin scale (SAM) was shown for 4 sec.

bimodal sessions, faces from the four Caucasian individuals were shown in each of the nine emotional facial expressions ranging from 100% happiness to 100% fear. In the unimodal auditory and bimodal sessions, each of the 18 sentences was pronounced once in happy and once in fearful intonation. In the bimodal sessions, faces were presented together with voices of the same gender and each of the nine emotional facial expressions of each of the four individuals was shown once together with a fearful and once in presence of a happy intonation. The sequence of bimodal and unimodal sessions was ordered randomly over subjects.

For the unimodal visual and auditory sessions, subjects were to rate the emotional valence of the presented stimulus (facial expression or emotional prosody, respectively). During bimodal stimulation, participants were instructed to judge the emotional valence of the facial expressions and to ignore the emotional content of the concomitant voice. All stimuli were judged on a nine-point SAM scale [Bradley and Lang, 1994], which was shown for 4 sec at 200 msec after stimulus offset (see Fig. 1). In the fMRI experiment, the subjects conveyed their decision via a fiberoptic system that allowed them to move a white dot on the SAM scale leftwards or rightwards by pressing the corresponding buttons in their right or left hand. To avoid lateralization effects caused by motor responses, the arrangement of the symbols

of the SAM scale was flipped in the horizontal direction for half of the participants of the fMRI experiment.

### Analysis of Behavioral Data

Valence ratings of facial expressions were submitted to a two-way analysis of variance (ANOVA) for repeated measures with facial expression (FE; nine steps) and affective prosody (AP; none, happy, and fearful) as within-subject factors. All resulting *P* values were corrected for heterogeneous correlations [Geisser and Greenhouse, 1958]. Separate post-hoc paired *t*-tests were used to compare valence ratings of facial expressions for each of the nine levels of emotional expression in the presence and absence of an emotionally spoken sentence. For graphical display of results, between-subject variability was removed according to the procedure described by Loftus and Masson [1994] to ensure that standard error bars exclusively represent within-subject variance.

### Image Acquisition

Functional MR images covering the whole brain (field of view [FOV] = 192 mm × 192 mm<sup>2</sup>, 28 slices, 4-mm slice thickness, and 1-mm gap) were acquired on a 1.5 T whole body scanner (Siemens VISION; Siemens, Erlangen, Germany) using an echo-planar imaging (EPI) sequence (repe-

tion time [TR] = 3 s, echo time [TE] = 39 ms, and matrix = 64<sup>2</sup>). Stimulus onset was jittered relative to the scan onset in steps of TR/4 resulting in intertrial intervals ranging from 9 to 15 s. The first five EPI images of each session were discarded to exclude measurements preceding T<sub>1</sub> equilibrium. High-resolution T<sub>1</sub>-weighted images were acquired using a magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (FOV = 256 mm, 192 slices, 1.5-mm slice thickness, no gap, flip angle 8 degrees, TR = 9.7 ms, TE = 4 ms, and matrix size = 256<sup>2</sup>).

### Image Analysis

Image analysis was carried out with *SPM2* software (Wellcome Department of Imaging Neuroscience, London, UK; <http://www.fil.ion.ucl.ac.uk/spm>). Preprocessing of functional MR images included movement correction, slice time correction, and coregistration with the anatomical data. The transformation matrix for normalization to Montreal Neurological Institute (MNI) space [Collins et al., 1994] was calculated based on the structural T<sub>1</sub>-weighted 3-D data set of each subject and subsequently applied to the functional images. Before statistical analysis, the functional MR images were smoothed using a Gaussian filter with 10-mm full-width half-maximum (FWHM). Statistical evaluation relied on a general linear model. Separate regressors were defined for each trial using a stick function convolved with the hemodynamic response function (HRF). For unimodal visual and bimodal sessions, events were time-locked to presentation of faces. For unimodal auditory presentation, stimulus onsets were set to 1 sec before the end of the sentence. To account for motion artifacts, realignment parameters were included in the model as covariates of no interest.

To determine which brain regions were involved in explicit and implicit processing of fearful and happy prosody, hemodynamic responses to the respective intonations were contrasted against each other both for the unimodal auditory sessions, in which the subjects explicitly rated the valence of the emotional prosody, and for the bimodal sessions, in which subjects were instructed to judge exclusively the facial expression.

To investigate neural responses during explicit judgment of facial expressions attributable to the emotional valence of these faces, BOLD responses to happy and fearful faces (ranging from 50% to 100% of the respective emotions) were contrasted with each other as well as with BOLD responses to neutral facial expressions (ranging from 25% happiness via absolutely neutral facial expressions to 25% fear) for both unimodal and bimodal presentation of faces.

To isolate brain regions that mediate effects of fearful prosody on cognitive evaluation of facial expressions, the contrast images representing the differences in BOLD responses to facial expressions in presence of a fearful voice as compared to the sole presentation of the same faces were calculated separately for each of the 36 visual stimuli. This was done to remove possible effects attributable to the physical properties and emotional valence of the visual stimuli. The resulting contrast images were correlated with the difference in valence

rating of facial expression between the two conditions, which was calculated for each facial expression and subject separately based on the behavioral data obtained during the fMRI experiment. Correlation coefficients were calculated from *t*-maps derived with standard statistical parametric mapping (SPM) algorithms by

$$r = \frac{t}{\sqrt{t^2 + (n-2)}} \quad (1)$$

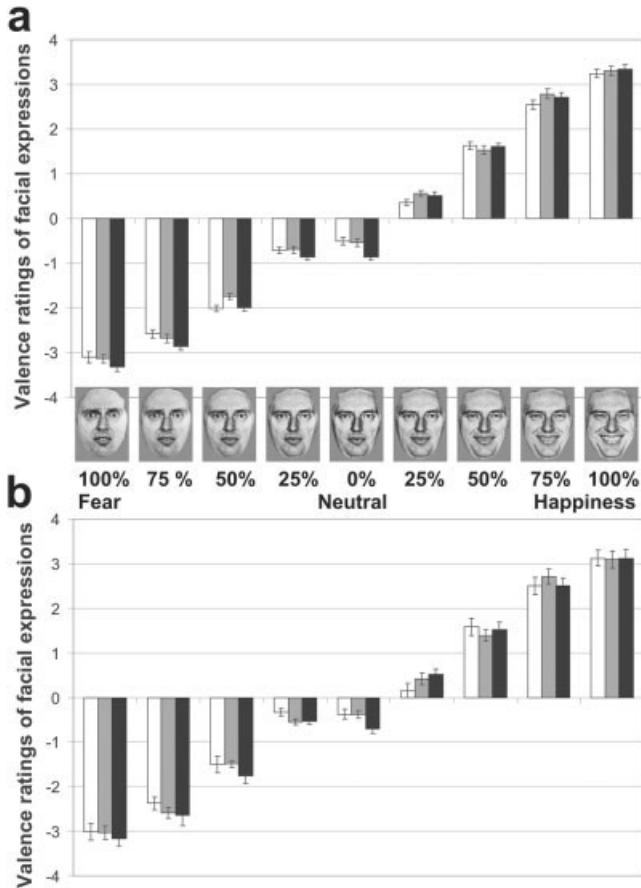
Resulting *r* values were Fisher Z-transformed and statistical evaluation of group data was based on a second-level random-effect analysis [Anders et al., 2004]. An analogous analysis was carried out for happy intonations.

Brain activations are reported on a height threshold of  $P < 0.001$  (uncorrected). Correction for multiple comparisons was carried out at cluster level (extent threshold:  $k = 43$ ,  $P < 0.05$ ). Activation clusters within the amygdala were assessed at a height threshold of  $P < 0.01$  (uncorrected). Small volume correction [SVC; Worsley et al., 1996] was carried out for a spherical volume (6-mm radius;  $P < 0.05$  corrected, a priori coordinates as derived from the Talairach and Tournoux [1988] atlas and transformed into MNI space:  $x = 21$ ;  $y = -3$ ;  $z = -18$  and  $x = -21$ ;  $y = -3$ ;  $z = -18$  for the right and left amygdala, respectively).

## RESULTS

### Behavioral Results

A two-way ANOVA for repeated measurements revealed a significant main effect for FE on valence rating ( $F[8,256] = 752.12$ ,  $P < 0.001$ ) in the behavioral experiment. The main effect of AP ( $F[2,64] = 8.95$ ,  $P < 0.01$ ) and the interaction FE  $\times$  AP ( $F[16,512] = 5.82$ ,  $P < 0.001$ ) were also significant, reflecting the fact that subjects rated fearful (25–100% fear) and neutral facial expressions as being more fearful when presented in presence of a fearfully spoken sentence as compared to that in the no-voice condition. Post-hoc paired *t*-tests showed that this effect was significant for all expression levels ranging from 100% fear to neutral (all paired  $T > 2.43$ ,  $P < 0.05$ ), whereas no significant shifts in interpretation occurred for happy (25–100% happiness) facial expressions (see Fig. 2a). The behavioral data collected during the fMRI experiment showed the same trend as that in the behavioral study (see Fig. 2b). Only the main effect of FE was significant ( $F[8,88] = 239.38$ ), however, whereas the effect of AP ( $F[2,22] = 1.93$ ,  $P = 0.18$ ) and the interaction FE  $\times$  AP ( $F[16,176] = 1.57$ ,  $P = 0.17$ ) failed to reach statistical significance due to the smaller sample size. Post-hoc paired *t*-tests showed that neutral facial expressions were rated as more fearful when presented together with a fearful voice than when presented without auditory stimulation (paired  $T[11] = 2.55$ ,  $P < 0.05$ ).



**Figure 2.**

Valence rating of facial expressions (mean  $\pm$  standard error) in presence of happy affective prosody (white), fearful affective prosody (black), and without concomitant acoustic stimulus (gray) obtained in the behavioral (a) and the fMRI experiment (b).

### Functional MRI Results

#### Explicit processing of emotional prosody

Contrasting brain activations elicited by happy and fearful intonations against each other in the unimodal auditory sessions did not reveal any clusters that survived correction for multiple comparisons for the whole brain. Even at a more liberal threshold of  $P < 0.01$ , neither fearful nor happy prosody resulted in a differential response within the amygdala.

#### Implicit processing of emotional prosody

The middle section of the right fusiform gyrus showed a significantly stronger hemodynamic response when facial expressions were shown in presence of a fearful voice as compared to that in the presence of a happy voice (MNI coordinates 24 -69 -15,  $Z = 3.91$ ,  $k = 46$ ; see Fig. 3).

Implicit processing of emotional prosody yielded stronger activation in the bilateral amygdalae ( $Z = 2.55$  and  $Z = 2.46$

for the right and left amygdala, respectively) for fearful intonations as compared to happy intonations. This difference failed to reach significance, however, within our region of interest.

#### Explicit processing of facial expressions

For both unimodal and bimodal presentation of faces, contrasting hemodynamic responses elicited by happy or fearful facial expressions against each other or to neutral facial expressions did not reveal any clusters that survived correction for multiple comparisons for the whole brain. Even at a more liberal threshold of  $P < 0.01$  (uncorrected), neither extrastriate cortices nor amygdalae showed any differential responses.

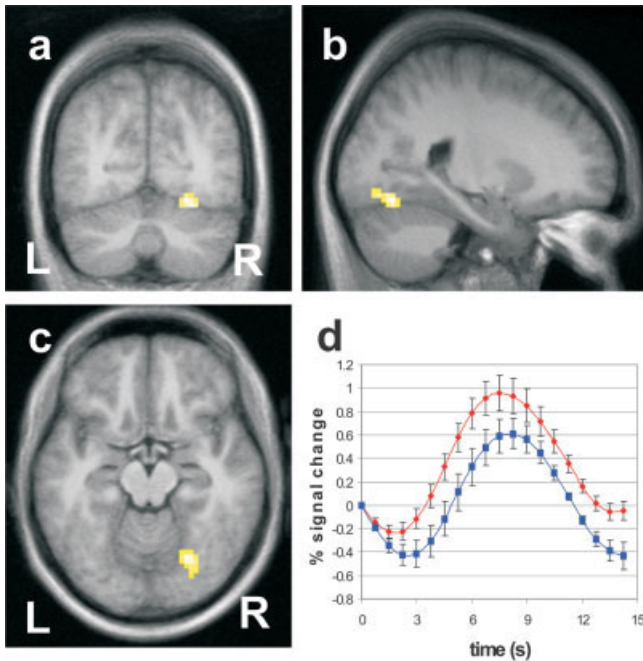
#### Modulation of explicit rating of facial expressions by emotional prosody

Regions showing a correlation between the difference in valence rating of facial expressions in presence of a fearful voice as compared to unimodal presentation of the same faces and the difference in BOLD-response of the two conditions included the basolateral part of the left amygdala extending into the periamygdaloid cortex (MNI coordinates -24 -6 -24,  $Z = 3.84$ ,  $k = 42$ ,  $P < 0.05$  SVC; see Fig. 4), the left postcentral gyrus (MNI coordinates -57 -18 36,  $Z = 3.40$ ,  $k = 11$ ), and the left middle temporal gyrus (MNI coordinates -48 -27 9,  $Z = 2.83$ ,  $k = 5$ ). The analogous correlation analysis for happy prosody did not reveal any suprathreshold clusters.

### DISCUSSION

In the behavioral experiment, fearfully spoken sentences were found to influence the valence rating of emotional facial expressions. Crossmodal interaction effects between auditory and visually presented nonverbal emotional information have been reported for the emotions happiness and sadness [de Gelder and Vroomen, 2000]. Our results differ from that study insofar as the behavioral effects in our study are due to congruency effects that only occurred in fear, whereas the interaction effects described previously [de Gelder and Vroomen, 2000] were attributable mostly to incongruity of audiovisual emotional information of both positive and negative valence. The longer duration of the visual presentation of the faces in our study (1,000 ms as compared to 350 ms) is the most probable reason for the lack of incongruity effects in our study. The fact that congruency effects on valence ratings of facial expressions occurred only in fear might be related to the higher biological relevance of a signal indicating threat or danger via two different sensory channels as compared to positive emotions, which are important in social interactions but of less immediate survival value.

Using the same paradigm as in the behavioral experiment in an event-related fMRI study, we demonstrated enhanced BOLD responses in the midsection of right fusiform gyrus



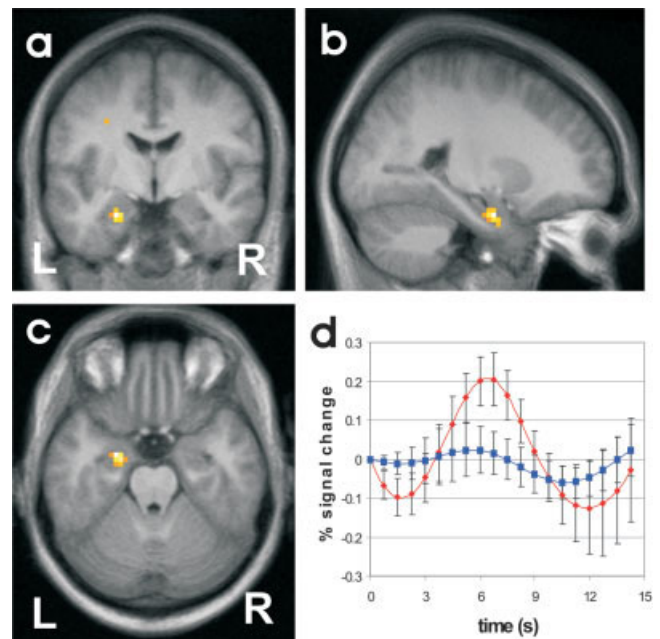
**Figure 3.**

Significantly stronger activation in the right fusiform gyrus during judgment of facial expressions in presence of a fearful voice as compared to a happy voice shown on coronal (a), sagittal (b), and transversal (c), slices ( $P < 0.001$ , uncorrected; correction at cluster level  $k > 43$ ,  $P < 0.05$ ). Event-related responses (d) in the right fusiform gyrus during judgment of faces in presence of fearful (red) and happy (blue) intonations.

when subjects rated the valence of facial expressions in presence of a fearfully spoken sentence as compared to that with happy intonations. Both lesion studies in prosopagnosic patients [Barton et al., 2002] and neuroimaging studies in normal subjects [Kanwisher et al., 1997; Puce et al., 1995] suggest that the midsection of the right fusiform gyrus is crucial for processing of faces. Furthermore, this extrastriate cortical region shows stronger responses to emotional as compared to neutral faces [Morris et al., 1998] and a preferential increase of neuronal activity to signals of danger has been demonstrated [Surguladze et al., 2003]. We extend these findings by showing that increases of neuronal responses in extrastriate brain regions during processing of facial expressions do not only occur if the face itself contains threat-related information, but can also be induced by a concomitant auditory signal that represents fear. The increased hemodynamic responses within the fusiform gyrus in presence of an auditory sign of threat might reflect enhanced vigilance for detection of fear in visual cues and an altered threshold for detection of such signals might induce shifts in interpretation of neutral and fearful facial expressions. No differential response in the extrastriate cortex or the amygdala to fearful or happy facial expressions was found when compared to that with neutral faces. This finding is in agreement with results from a previous neuroim-

aging study demonstrating that fearful facial expressions activate the right fusiform gyrus stronger than neutral faces do during passive viewing or gender-differentiation tasks, but not during explicit rating of emotional intensity [Lange et al., 2003]. This suggests that task instructions modulate neural responses to facial expressions. Similarly, attenuation of amygdala responses during explicit emotion labeling of facial expressions has been reported [Critchley et al., 2000; Hariri et al., 2000, 2003].

Although there is converging evidence from both lesion [Adolphs et al., 1994; Calder et al., 1996] and neuroimaging studies [Breiter et al., 1996; Morris et al., 1996] arguing for a specific contribution of the amygdalae in perception of fearful facial expressions, their role in processing of fearful prosody is less clear [for review see Adolphs et al., 2002]. In one patient with bilateral lesions of the amygdala, impaired recognition of fearful emotional vocalizations was found [Scott et al., 1997]; however, this finding could not be replicated by others [Adolphs and Tranel, 1999; Anderson and Phelps, 1998]. Although neuroimaging studies that examined neuronal responses during implicit processing of emotional vocalizations revealed differential responses within the right amygdala [Morris et al., 1999; Phillips et al., 1998], studies using explicit emotion-labeling paradigms failed to demonstrate valence effects to emotional vocalizations



**Figure 4.**

Correlation analysis between the crossmodal impact of fearful voices on judgment of facial expressions and BOLD response in the left amygdala on coronal (a), sagittal (b), and transversal (c) slices ( $P < 0.01$  uncorrected;  $P < 0.05$ , small volume corrected). Event-related responses (d) of trials in which faces were judged as being more fearful when presented together with a fearful voice (red) and trials where no such shift in interpretation occurred (blue).

[Royet et al., 2000] and emotional prosody [Wildgruber et al., 2002, 2005]. In view of the observation that task instructions are crucial for activations to emotional facial expressions within the amygdalae [Critchley et al., 2000] and the hypothesis that the nature of the task might be also important for amygdalar activation during processing of fearful intonations, we contrasted BOLD responses to fearful and happy intonations for both explicit and implicit processing of emotional intonations. Neither implicit nor explicit processing revealed significant valence effects within the amygdala. A considerable degree of caution has to be exercised with this negative result, however, especially because sub-threshold activation was observed in the bilateral amygdala during implicit processing of fearful as compared to happy intonations.

As its main goal, the present study was designed to clarify whether the amygdala mediates shifts in interpretation of emotional facial expressions in presence of a fearful voice. To this end, we calculated the correlation between the BOLD response induced by a fearful intonation and its effect on the behavioral level. In accordance with our predictions, a significant correlation was found in the left amygdala. This is consistent with the view that the amygdala modulates neuronal activity in brain regions concerned with visual processing [Davis and Whalen, 2001; Morris et al., 1998; Vuilleumier et al., 2004] and findings of a recent fMRI study suggesting that the left amygdala integrates audiovisual fear-related emotional information to a common percept [Dolan et al., 2001]. Assignment of specific functions to different substructures of the amygdala is difficult due to low spatial resolution of the functional images ( $3 \times 3 \times 4$  mm) acquired in this study and introduction of a smoothing kernel with 10-mm FWHM. The cluster showing this correlation, however, was found in the basolateral part of the amygdala extending into the adjacent periamygdaloid cortex. Activations in the basal parts of the amygdala in paradigms involving fear conditioning or visual stimuli showing fear is a common observation across studies [Büchel et al., 1998; Dolan et al., 2001; Morris et al., 1996] and in one of these studies the strongest evidence for activation was found in the periamygdaloid cortex [Morris et al., 1996]. Furthermore, anatomical data obtained from rodent brains indicate that projections from cortices conveying sensory-related information converge in the basolateral amygdala complex and in the periamygdaloid cortex [McDonald, 1998; Pitkänen, 2000]. In the lateral nucleus, these projections are organized topographically with the axons from the visual and auditory cortices terminating in more caudal portions than the projections from gustatory, visceral, and somatosensory cortices [McDonald, 1998]. The finding of a significant correlation in the basolateral part of the amygdala therefore is in agreement with the suggestion that integration of sensory information obtained from different sensory channels is mediated via the basolateral amygdala complex and the periamygdaloid cortex [Pitkänen, 2000], but needs confirmation by fMRI experiments with a higher spatial resolution. Other regions that showed a correlation between

BOLD response and behavioral effect included the left post-central gyrus and left middle temporal gyrus. The middle temporal gyrus was described recently as being involved in audiovisual integration of nonverbal affective information [Pourtois et al., 2005]. The finding of a correlation in the left middle temporal gyrus in our study can only be reported descriptively, however, because no a priori hypothesis was made for this region.

An activation pattern including the left amygdala and right fusiform gyrus during processing of fear in voice and face has been described [Dolan et al., 2001]. The differential roles of these structures in integration of audiovisual affective information, however, could not be disambiguated in that study. Based on our findings, we suggest that increased neuronal activity in face-responsive regions of the extrastriate cortex in the presence of a fearful voice corresponds to enhanced alertness to facial signs of threat when other channels signal danger, whereas responses in the left amygdala are correlated with the behavioral output.

## CONCLUSIONS

Fearful prosody was found to influence judgment of emotional facial expressions. In presence of fearful intonations, neutral and fearful facial expressions were perceived as being more fearful as compared to facial expressions without concomitant auditory stimulus. A significant correlation between the shift in interpretation of facial expressions in presence of a fearful voice and the BOLD response was found in left basolateral amygdala, suggesting that the impact of voice on processing of faces is mediated via this neuronal structure. Moreover, increased BOLD responses within the mid-portion of the right fusiform gyrus were found during judgment of facial expressions in presence of a fearful as compared to a happy voice. These augmented responses in the extrastriate visual cortex might reflect enhanced vigilance to facial signs of fear in the presence of auditory signals representing menace or danger.

## ACKNOWLEDGMENTS

This study was supported by the Deutsche Forschungsgemeinschaft (grant WI 2101/1-1) and by the Junior Science program of the Heidelberger Academy of Sciences and Humanities.

## REFERENCES

- Adolphs R (2002): Neural systems of recognizing emotion. *Curr Opin Neurobiol* 12:169–177.
- Adolphs R, Tranel D, Damasio H, Damasio A (1994): Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372:669–672.
- Adolphs R, Tranel D (1999): Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia* 37:1285–1292.
- Anders S, Lotze M, Erb M, Grodd W, Birbaumer N (2004): Brain activity underlying emotional valence and arousal. *Hum Brain Mapp* 23:200–209.

- Anderson AK, Phelps EA (1998): Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. *Neuroreport* 9:3607–3613.
- Barton JJ, Press DZ, Keenan JP, O'Connor M (2002): Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology* 58:71–78.
- Bradley MM, Lang PJ (1994): Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J Behav Ther Exp Psychiatry* 25:49–59.
- Breiter HC, Etkoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, Strauss MM, Hyman SE, Rosen BR (1996): Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 2:875–887.
- Büchel C, Morris JS, Dolan RJ, Friston KJ (1998): Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* 20:947–957.
- Calder AJ, Young AW, Rowland D, Perrett DI, Hodges JR, Etkoff NL (1996): Face perception after bilateral amygdala damage: differentially severe impairment of fear. *Cognit Neuropsychol* 13:699–745.
- Collins DL, Neelin P, Peters TM, Evans AC (1994): Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 18:192–205.
- Critchley H, Daly E, Phillips M, Brammer M, Bullmore E, Williams S, Van Amelsvoort T, Robertson D, David A, Murphy D (2000): Explicit and implicit neural mechanisms for processing of social information from facial expressions: a functional magnetic resonance imaging study. *Hum Brain Mapp* 9:93–105.
- Davis M, Whalen PJ (2001): The amygdala: vigilance and emotion. *Mol Psychiatry* 6:13–34.
- de Gelder B, Vroomen J (2000): The perception of emotions by ear and eye. *Cognit Emotion* 14:289–311.
- Dolan RJ, Morris JS, de Gelder B (2001): Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci USA* 98:10006–10010.
- Ekman P, Friesen WV. 1976. *Pictures of facial affect*. Palo Alto: Consulting Psychologists Press.
- Freese JL, Amaral DG (2005): The organization of projections from the amygdala to visual cortical areas TE and V1 in the macaque monkey. *J Comp Neurol* 486:295–317.
- Geisser S, Greenhouse SW (1958): An extension of Box's results on the use of the F-distribution in multivariate analysis. *Ann Math Stat* 29:885–891.
- Hariri AR, Bookheimer SY, Mazziotta JC (2000): Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport* 11:43–48.
- Hariri AR, Mattay VS, Tessitore A, Fera F, Weinberger DR (2003): Neocortical modulation of the amygdala response to fearful stimuli. *Biol Psychiatry* 52:494–501.
- Kanwisher N, McDermott J, Chun MM (1997): The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Lang PJ (1995): The emotion probe. Studies of motivation and attention. *Am Psychol* 50:372–385.
- Lange K, Williams LM, Young AW, Bullmore ET, Brammer MJ, Williams SCR, Gray JA, Phillips ML (2003): Task instructions modulate neural responses to fearful facial expressions. *Biol Psychiatry* 53:226–232.
- Loftus GR, Masson ME (1994): Using confidence intervals in within-subjects designs. *Psychon Bull Rev* 1:476–490.
- Massaro DW, Egan PB (1996): Perceiving affect from the voice and the face. *Psychon Bull Rev* 3:215–221.
- McDonald AJ (1998): Cortical pathways to the mammalian amygdala. *Prog Neurobiol* 55:257–332.
- Morris JS, Frith CD, Perrett DI, Rowland D, Young AW, Calder AJ, Dolan RJ (1996): A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383:812–815.
- Morris JS, Friston KJ, Büchel C, Frith CD, Young AW, Calder AJ, Dolan RJ (1998): A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain* 121:47–57.
- Morris JS, Scott SK, Dolan RJ (1999): Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia* 37:1155–1163.
- Perrett I, May KA, Yoshikawa S (1994): Facial shape and judgements of female attractiveness. *Nature* 368:239–42.
- Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, Williams SC, Bullmore ET, Brammer M, Gray JA (1998): Neural responses to facial and vocal expressions of fear and disgust. *Proc Biol Sci* 265:1809–1817.
- Pitkänen A. 2000. Connectivity of the rat amygdaloid complex. In: Aggleton J, editor. *The amygdala: a functional analysis*. New York: Oxford University Press. p 31–116.
- Pourtois G, de Gelder B, Bol A, Crommelinck M (2005): Perception of facial expressions and voices and of their combination in the human brain. *Cortex* 41:49–59.
- Puce A, Allison T, Gore J, McCarthy G (1995): Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol* 74:1192–1199.
- Royet JP, Zald D, Versace R, Costes N, Lavenne F, Koenig O, Gervais R (2000): Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. *J Neurosci* 20:7752–7759.
- Scott SK, Young AW, Calder AJ, Hellowell DJ, Aggleton JP, Johnson M (1997): Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385:254–257.
- Surguladze SA, Brammer MJ, Young AW, Andrew C, Travis MJ, Williams SC, Phillips ML (2003): A preferential increase in the extrastriate response to signals of danger. *Neuroimage* 19:1317–1328.
- Talairach J, Tournoux P. 1988. *Co-planar stereotactic atlas of the human brain*. Stuttgart: Thieme.
- Vroomen J, Driver J, de Gelder B (2001): Is cross-modal integration of emotional expressions independent of attentional resources? *Cogn Affect Behav Neurosci* 1:382–387.
- Vuilleumier P, Armony JL, Driver J, Dolan RJ (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 MP, Armony JL, Driver J, Dolan RJ (2004): Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nat Neurosci* 7:1271–1278.
- Wildgruber D, Pihan H, Ackermann H, Erb M, Grodd W (2002): Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage* 15:856–869.
- Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H (2005): Identification of emotional intonation evaluated by fMRI. *Neuroimage* 24:1233–1241.
- Worsley K, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans A (1996): A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4:74–90.