Rapid influence of emotional scenes on encoding of facial expressions: an ERP study

Ruthger Righart,1 and Beatrice de Gelder1,2
1Cognitive and Affective Neurosciences Laboratory, Tilburg University, The Netherlands and 2Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Bldg. 36, Main St., Charlestown, MA 02129, USA

In daily life, we perceive a person’s facial reaction as part of the natural environment surrounding it. Because most studies have investigated how facial expressions are recognized by using isolated faces, it is unclear what role the context plays. Although it has been observed that the N170 for facial expressions is modulated by the emotional context, it was not clear whether individuals use context information on this stage of processing to discriminate between facial expressions. The aim of the present study was to investigate how the early stages of face processing are affected by emotional scenes when explicit categorizations of fearful and happy facial expressions are made. Emotion effects were found for the N170, with larger amplitudes for faces in fearful scenes as compared to faces in happy and neutral scenes. Critically, N170 amplitudes were significantly increased for fearful faces in fearful scenes as compared to fearful faces in happy scenes and expressed in left-occipito-temporal scalp topography differences. Our results show that the information provided by the facial expression is combined with the scene context during the early stages of face processing.

Keywords: context; face perception; emotion; event-related potentials; P1; N170

INTRODUCTION

The recognition of facial expressions has traditionally been studied by using isolated faces (Ekman, 1992; Adolphs, 2002). However, facial expressions can be rather ambiguous when viewed in isolation. This ambiguity may be resolved if the accompanying context is known (de Gelder et al., 2006; Barrett et al., 2007). In comparison to the study of object perception (Palmer, 1975; Bar, 2004; Davenport and Potter, 2004), there have only been a few behavioral studies that investigated the question how contexts may influence face processing. Facial expressions of fear tend to be perceived more frequently as expressing anger when subjects had heard a story about an anger provoking situation in advance (e.g. about people who were rejected in a restaurant) (Carroll and Russell, 1996). Facial expressions that were viewed in the context of emotional scenes were categorized faster in congruent visual scenes (e.g. faster recognition of a face conveying disgust in front of a garbage area) than in incongruent scenes (Righart and de Gelder, 2008). fMRI studies have also shown that facial expressions are interpreted differently given the context information that is available (Kim et al., 2004; Mobbs et al., 2006).

Behavioral and fMRI studies are not able to show at what stage of processing emotional contexts affect face recognition. This may relate to an early stage of encoding, but it may also relate to a later stage of semantic associations that are made between face and context. Because of the time-sensitivity of event-related potentials (ERPs), it is possible to investigate how contexts affect different stages of face processing. The N170 is an ERP component that has been related to face encoding (Bentin et al., 1996; George et al., 1996; Itier and Taylor, 2004). The N170 occurs at around 170 ms after stimulus onset and has a maximal negative peak on occipito-temporal sites.

Although some studies have observed that the N170 is insensitive to facial expressions (Eimer and Holmes, 2002; Holmes et al., 2003), other studies have shown that the N170 amplitude is modified by facial expressions of emotion (Batty and Taylor, 2003), especially for fearful expressions (Batty and Taylor, 2003; Stekelenburg and de Gelder, 2004; Caharel et al., 2005; Righart and de Gelder, 2006; Williams et al., 2006). This suggests that emotional information may affect early stages of face encoding.

Another ERP component, the P1, with a positive deflection occurring at occipital sites at around 100 ms after stimulus onset, has also been related to face processing. Most studies have centred on this component because of its relation to spatial attention and physical features (Hillyard and Anllo-Vento, 1998). However, recent studies have found that the P1 amplitude is larger for faces than for nonface objects (Itier and Taylor, 2004; Herrmann et al., 2005), and that facial expressions affect the P1 amplitude as well (Batty and Taylor, 2003; Eger et al., 2003). The P1 may reflect a stage of face detection, and precedes the N170 that may reflect a stage of configural processing (Itier and Taylor, 2004). This time-course is consistent with earlier suggestions that global processing of faces occurs at around 117 ms,
while fine processing of facial identity and expressions may occur at around 165 ms (Sugase et al., 1999).

An important question is how emotional contexts in which faces are perceived affect the aforementioned P1 and N170 components. In our previous study it was found that the N170 amplitude was larger when the face appeared in a fearful scene, especially when the face expressed fear (Righart and de Gelder, 2006). This effect was not found for the P1, though effects have been observed on this component for scenes only (Smith et al., 2003; Carretie et al., 2004; Righart and de Gelder, 2006), and for facial expressions in interaction with bodily expressions (Meeren et al., 2005).

While previous studies have reported that the N170 elicited by facial expressions is modulated by the emotional context, it is not yet clear whether individuals use context information in this stage when the task requires them to attend to the face and to discriminate explicitly between facial expressions. We hypothesized that the N170 component will still be affected by emotion regardless of the changed task conditions (Caharel et al., 2005).

In addition, it was not clear from our previous study how facial expressions of fear are processed compared with other facial expressions (e.g. happiness) as a function of the emotional context. In the present study, we investigated whether explicit categorization of facial expressions (fear, happiness) affects the P1 and the N170 for faces. As the discrimination of fine expressions may be associated with later processing stages than the P1 (compare with Sugase et al., 1999), and especially facial expressions of fear affect the N170 amplitude, we hypothesized that the N170 amplitude would be larger for fearful faces in fearful scenes as compared to happy and neutral scenes.

**MATERIALS AND METHODS**

**Subjects**

Eighteen participants (12 male) ranging from 21 to 50 years participated in the experiment (13 right handed). One participant was removed because of visual problems (glaucoma). The remaining participants had normal or corrected-to-normal vision. None of them reported a history of neurological or psychiatric diseases. All participants had given informed consent and were paid €20 for participation. The study was performed in accordance with the ethical standards of the Declaration of Helsinki.

**Stimuli**

Stimuli consisted of faces that were centrally overlaid on pictures of natural scenes (Figure 1). Stimuli were color pictures (24 male and female) of facial expressions of fear and happiness from the Karolinska Directed Emotional Faces set (Lundqvist et al., 1998) and color pictures of natural scenes selected from the International Affective Picture System (IAPS) (Lang et al., 1999, p. 7560, 9622, 9911, 9920), which were complemented with images from the web to have for each category an equal number of scenes (Figure 1). Contents were validated for emotions of fear, happiness or neutrality and arousal (seven-point scale,
from 1 = unaffected to 7 = extremely aroused) by a different
group of participants (N=10) using a collection of 401
picture stimuli. In total, 24 scene stimuli were selected for
each emotion category (Figure 1). For the final set that was
used in the experiment, arousal rates for fearful (4.84) and
happy scenes (4.24) were significantly higher than neutral
scenes (2.00) (both P < 0.001 but P > 0.05 for fear–happy).
On average, the intended label for happy scenes was chosen
for 75% of the trials, for fearful scenes for 64% of the trials
and for neutral scenes for 87% of the trials.

In order to control for low-level features (e.g. color),
all scene pictures were scrambled by randomizing the position
of pixels across the image (blocks of 8 x 8 pixels were randomized across the image measuring 768 x 572 pixels
width and height), which makes the pictures meaningless.
The resulting pictures were inspected carefully for residual
features that could cue recognition.

The height and width of the facial images was
7.9 cm x 5.9 cm (5.6 x 4.2”) and for context images
24.5 cm x 32 cm, respectively (17.4” x 22.6”). Faces did not
occlude critical information in the context picture.

**Design and procedure**

Participants were seated in a dimly illuminated and an
electrically shielded cabin with a monitor positioned at
80 cm distance. Participants were instructed and familiarized
with the experiment by a practice session.

The experiment was run in 16 blocks each containing
72 trials of face–context compounds, 8 blocks of faces with
intact context scenes and 8 blocks of faces with scrambled
scenes. Blocks presenting intact and scrambled images alternated
(with order randomized across participants). Facial
expressions (fear, happy) were paired with a scene of each
category (fear, happy, neutral) to have a balanced factorial
design. Categories of emotional scenes appeared randomly
throughout each block, so that each condition was presented
in each block. Altogether, intact and scrambled blocks
amounted to 12 conditions of each 96 trials.

The face-context compounds were presented for 200 ms
and were preceded by a fixation cross. Participants performed a two-alternative forced choice task in which they
categorized facial expression as happy or fearful. Responses
were recorded from stimulus onset. No feedback was given.
They were instructed to respond as accurately and fast as possible. Response buttons were counterbalanced across participants. The intertrial interval was randomized between
1200 and 1600 ms.

After the main experiment, participants received a validation task in which they judged the arousability of the scenes and categorized them by emotion. In contrast to the stimulus validation, arousal was measured on a five-point scale
now because of response box limitations in the EEG experiment (five-point scale from 1 = calm to 5 = extremely arousing). Emotional categorization was measured by using three options (fear, happy, neutral).

**EEG recording**

EEG was recorded from 49 active Ag–AgCl electrodes
(BioSemi Active2) mounted in an elastic cap referenced
to an additional active electrode (Common Mode Sense).
EEG was bandpass filtered (0.1–30 Hz, 24 dB/Octave). The
sampling rate was 512 Hz. All electrodes were referenced
offline to an average reference. Horizontal electrooculogra-
phies (hEOG) and vertical electrooculographies (vEOG)
were recorded. The raw data were segmented into epochs
from 200 ms before and 1000 ms after stimulus onset. The
data were baseline corrected to the first 200 ms.

EEG was EOG corrected by using the algorithm of Gratton
et al. (1983). Epochs exceeding 100 μV amplitude difference
at any channel were removed from analyses. No differences
were observed across conditions for facial expressions and
emotional scenes. On average 87.5 trials (range across con-
ditions: 86.8–88.5) were left for faces in intact scenes and
85.1 trials (84.4–85.4) were left for faces in scrambled scenes
after removal of the artifacts. After removal of trials containing
inaccurate responses or responses below 200 ms, ERPs
were averaged for conditions of facial expressions of fear
in intact fearful, happy and neutral scenes, and for happy
facial expressions in the same context categories. Similarly,
averages were computed for scrambled blocks, resulting in a
total of 12 conditions.

Electrode selection for P1 and N170 analyses was based on
previous studies. Based on grand average ERP inspection,
peak detection windows for P1 and N170 were measured using time-windows of 60–140 ms and 100–220 ms, respectively.
Peak latencies and amplitudes of P1 were analyzed at occipital sites (O1/2) and occipito-temporal sites (PO3/4,
PO7/8) as the maximal positive peak amplitude. The N170
was analyzed on occipito-temporal sites (P5/6, P7/8 and
PO7/8) as the maximal negative peak amplitude.

**Data analyses**

Behavioral analyses were performed for error rate (percentage of incorrect responses) and RTs (average response-times for correct responded trials). RT data were inspected for outliers for each participant. RTs > 2.5 SD from the mean of each condition were removed from analyses. Using these
criteria, 2.6% of the trials were removed. Main and interaction
effects were analyzed by using repeated measures ANOVA containing the factors Image (intact, scrambled),
Facial expression (fear, happy) and Scene (fear, happy, neutral). Planned comparisons were performed (Howell, 2002)
to test our specific hypothesis that fearful faces are faster
recognized in fearful scenes (congruent) as compared to
happy scenes (incongruent), and that happy faces are faster
recognized in happy scenes (congruent) as compared to
fearful scenes (incongruent). (α = 0.05, one-tailed t-test, directional hypotheses).

P1 and N170 latencies and amplitudes were analyzed by using repeated measures ANOVA containing the within subject factors image (intact, scrambled), facial expression
Scalp topographic analyses and differences across the topographies for emotional scenes were analyzed by using repeated measures ANOVA. Mean amplitudes around the N170/Vertex Positive Potential (VPP) (140–160 ms) were calculated and differences across topographies were tested by t-tests on each electrode site (Rousselet et al., 2004). Amplitudes were vector normalized according to the method employed by McCarthy and Wood (1985). P-values were corrected by Greenhouse–Geisser epsilon when appropriate. Statistics are indicated with original degrees of freedom (Picton et al., 2000).

RESULTS

Behavioral results

The average error rate across all conditions was below 5%. Because of this low rate, no planned comparisons were performed. The main effect for image was significant, $F(1,16) = 4.08, P < 0.05$, in that errors were slightly increased for faces in intact scenes ($M = 4.4\%$) than scrambled scenes ($M = 4.0\%$).

The analyses for RTs showed a main effect for facial expression, $F(1,16) = 4.07, P < 0.05$ as reflected in faster RTs to happy facial expressions ($M = 665 \text{ ms}$) than fearful facial expressions ($M = 690 \text{ ms}$), which is shown in Figure 2. A three-way interaction was observed between facial expression, scene and image, $F(2,15) = 3.89, P < 0.05$. Planned comparisons showed that happy faces were recognized faster in intact happy ($M = 667 \text{ ms}$) and neutral ($M = 668 \text{ ms}$) than fearful scenes ($M = 682 \text{ ms}$), $t(16) = 2.54, P < 0.05$ and $t(16) = 2.16, P < 0.05$, respectively. The differences were not significant between happy faces in scrambled versions of happy ($M = 661 \text{ ms}$), neutral ($M = 656 \text{ ms}$) and fearful scenes ($M = 657 \text{ ms}$). Neither were the differences significant for fearful faces in intact fearful scenes ($M = 691 \text{ ms}$), happy ($M = 694 \text{ ms}$) and neutral scenes ($M = 694 \text{ ms}$).

In the evaluation task after the main experiment, a main effect was found for arousal, $F(2,32) = 59.41, P < 0.001$, in that fearful ($M = 3.81$) and happy scenes ($M = 3.45$) were rated as being more arousing than the neutral scenes ($M = 1.84$), $t(16) = 11.99, P < 0.001$ and $t(16) = 8.80, P < 0.001$. No effects were found for the categorization task, $F(2,32) = 2.22, P > 0.05$. However, post-hoc t-tests showed that the ratings for fearful and happy categories differed significantly, $t(16) = 2.15, P < 0.05$. The intended label for happy scenes was chosen on 87% of the trials, for fearful scenes on 77% of the trials and for neutral scenes on 81% of the trials. Thus, these results are consistent with our previous ratings on arousal and category.

ERPs for faces in context

P1 component. The ERPs of 14 participants showed a distinctive P1 deflection. No main effects were observed for facial expression and scene for P1 latency and amplitude, $P > 0.05$. A marginally significant three-way interaction was found between image, electrode position and scene, $F(4,10) = 2.92, P = 0.08$, which was explained by amplitudes that were smaller for faces in intact happy ($M = 6.18 \mu V$) than intact fearful ($M = 6.97 \mu V$) and intact neutral scenes ($M = 6.91 \mu V$), but on electrode pair O1/2 only, respectively, $P < 0.01$ and $P < 0.05$. None of the comparisons for scrambled scenes were significant.

N170 component. The ERPs of 15 participants showed a distinctive N170. Main effects were found for the N170 latency of facial expression, $F(1,14) = 20.41, P < 0.001$ and scene, $F(2,13) = 15.47, P < 0.001$, but the effects were explained by slight latency differences between happy faces ($M = 148 \text{ ms}$) and fearful faces ($M = 149 \text{ ms}$) and between neutral ($M = 148 \text{ ms}$) and fearful scenes ($M = 149 \text{ ms}$), and therefore these effects are not further discussed.
The analyses for the N170 amplitude showed a main effect for facial expression, $F(1,14) = 6.36, P < 0.05$, in that N170 amplitudes were more negative to facial expressions of happiness ($M = -4.64 \mu V$) than fear ($M = -4.25 \mu V$).

A three-way interaction was observed between image, electrode-position and scene, $F(4,11) = 6.90, P < 0.05$ (Figure 3). Amplitudes were more negative for faces in intact fearful ($M = -4.51 \mu V$) than faces in intact happy ($M = -3.94 \mu V$) and intact neutral scenes ($M = -4.05 \mu V$), $P < 0.01$, $P < 0.05$, respectively. These results were significant on electrodes P7/8 only. It is unlikely that these effects are based on low-level features. The differences between faces in scrambled fearful ($M = -5.23 \mu V$) and faces in scrambled happy scenes ($M = -5.03 \mu V$) obtained marginal significance ($P = 0.06$). The differences between faces in scrambled fearful scenes ($M = -5.23 \mu V$) and faces in scrambled neutral scenes ($M = -5.26 \mu V$) were not significant.

Planned comparisons were performed on N170 peak amplitudes for fearful and happy facial expressions separately as a function of scene. It was found that the N170 amplitudes were more negative for fearful facial expressions in fearful scenes than facial expressions in happy scenes for P7 ($-4.18 \mu V; -3.19 \mu V$), $t(14) = 3.74, P < 0.01$, P8 ($-4.84 \mu V; -4.23 \mu V$), $t(14) = 2.59, P < 0.05$ and compared to neutral scenes for P7 ($-4.18 \mu V; -3.33 \mu V$), $t(14) = 3.06, P < 0.01$ (Figure 3). For happy facial expressions, N170 amplitudes were more negative in fearful scenes than happy facial expressions in happy scenes for P7 ($-3.90 \mu V; -3.52 \mu V$), $t(14) = 2.61, P < 0.05$, and happy facial expressions in neutral scenes for P7 ($-3.90 \mu V; -3.39 \mu V$), $t(14) = 2.28, P < 0.05$. The comparisons for other electrodes on the occipito-temporal scalp are shown in Figure 4. Importantly, the differences were obtained on all occipito-temporal electrodes on the left hemisphere for fearful faces in fearful scenes as compared to happy and neutral scenes, while this was not the case for happy faces. T-tests on all electrodes for the calculated mean amplitudes (140–160 ms) were similar to peak analyses (Figure 4).

Fig. 3 Grand-averaged ERPs of the N170/VPP components. N170/VPP for fearful (A) and happy faces (B) in fearful, happy and neutral scenes. The N170 is displayed for occipito-temporal electrode sites (P7/8) and the VPP is displayed for the vertex electrode (Cz). Negative amplitudes are plotted upwards.
Scalp topography analyses. Scalp topography analyses were performed to examine further whether the implied hemispheric differences for fearful facial expressions in fearful scenes were reflected in topographical differences. For the scalp topographic distribution, mean amplitudes were calculated around the N170 peak, which occurred on C24 150 ms. Figure 4 shows planned comparisons for mean amplitudes at 140–160 ms indicating that fearful faces were larger in fearful scenes than happy and neutral scenes on left occipito-temporal electrodes. In addition, the central electrodes show a positivity (associated with the VPP, Joyce and Rossion, 2005) that is increased for fearful faces in fearful scenes as compared to happy and neutral scenes. The topographic interaction between a large range of symmetrically positioned occipito-temporal electrodes (T7/8; TP7/8; CP3/4; CP5/6; P7/8; P5/6; PO7/8; PO3/4; O1/2) and scene was tested for facial expressions of fear and happiness separately by repeated measures ANOVA.

A significant topography difference was found between fearful facial expressions in intact fearful scenes as compared to fearful facial expressions in intact happy scenes, F(19,304) = 4.10, P < 0.05, but it was not different from fearful facial expressions in neutral scenes, F(19,304) = 1.28, P = 0.29. The difference between fearful facial expressions in happy scenes and fearful facial expressions in neutral scenes was significant neither, F(19,304) = 1.61, P = 0.18. The comparisons across scrambled scenes were also nonsignificant. For happy facial expressions, no topography differences were found between faces that were presented in fearful scenes compared to happy, F(19,304) = 1.92,
P = 0.13, and neutral scenes, $F(19,304) = 0.51$, $P = 0.74$, and nor between happy and neutral scenes, $F(19,304) = 0.93$, $P = 0.44$.

**DISCUSSION**

We showed that affective information provided by natural scenes influences face processing. Larger N170 amplitudes were observed for faces in fearful scenes as compared to faces in happy and neutral scenes, which were particularly increased for fearful faces on the left hemisphere electrodes. Taken together, our results indicate that information from task-irrelevant scenes is combined rapidly with the information from facial expressions, even when the task requires categorization of facial expressions. These results suggest that subjects use context information on this stage of processing when they discriminate between facial expressions. The emotional context may facilitate this categorization process by constraining perceptual choices (Barrett et al., 2007).

The behavioral data confirm earlier findings that context information may influence how facial expressions are recognized (Carroll and Russell, 1996; reviewed by Barrett et al., 2007). Categorization of happy facial expressions is significantly faster in happy than fearful scenes, whereas the faster recognition of fearful facial expressions in fearful scenes compared to happy scenes did not reach significance. These behavioral data replicate our previous findings (Righart and de Gelder, 2008), but appear to oppose the results on the ERP data. The differences for scrambled scenes were not significant and therefore it is unlikely that low-level features explain these results. A novel finding in this study is that RTs for happy faces in neutral scenes were also faster than fearful scenes. Future studies may investigate how recognition of facial expressions is affected by emotional scenes when individuals are not constrained by choosing between emotion labels (Barrett et al., 2007).

The behavioral data do not necessarily reflect the same brain processes as measured by ERPs. Behavioral results differ across tasks that are employed, as it has been reported that response times are fastest to negative expressions if participants had to detect expressions (Öhman et al., 2001), but are slowest if participants had to discriminate among several expressions (Calder et al., 2000). The ERP data showed different patterns for the P1 component compared to the N170 component. The P1 amplitudes were larger to faces in fearful than happy scenes, which is consistent with literature that has shown P1 effects for emotional scenes without faces (Smith et al., 2003; Carretie et al., 2004; Righart and de Gelder, 2006), and may relate to attentional effects that have been reported for the P1 (Hillyard and Anllo-Vento, 1998). However, it should be noted that the results for faces in neutral scenes were not consistent with this interpretation, because the amplitudes for neutral scenes were similar to these for fearful scenes. We cannot exclude here that low-level features may have introduced this effect for neutral scenes (Allison et al., 1993). However, low-level features are unlikely explaining the main results for the N170, because these results replicate our previous data but now by using different stimulus sets (Righart and de Gelder, 2006).

In contrast to the P1, increased N170 amplitudes were observed on the left hemispheric electrode sites when fearful faces were accompanied by fearful scenes rather than happy or neutral scenes. As fearful scenes increased the N170 amplitude for fearful faces significantly, emotions may combine specifically for fear at this stage of encoding. Previous work has already shown that the N170 amplitude is modified by facial expressions (Batty and Taylor, 2003), especially for fearful facial expressions (Batty and Taylor, 2003; Stekelenburg and de Gelder, 2004; Williams et al., 2006), and that fearful scenes increase the N170 amplitude for fearful facial expressions even more (Righart and de Gelder, 2006). Although the N170 amplitude was increased for both fearful and happy faces in fearful as compared to happy and neutral scenes, which may relate to general effects of arousal, specific effects were found for fearful faces on the left hemisphere, as indicated by the scalp topographic analyses that showed significant differences on left occipito-temporal electrodes.

The left hemispheric effects may correspond to contextual effects that were observed before in an fMRI study (Kim et al., 2004). It was found that left fusiform gyrus activation to surprised faces was increased if preceded by a negative context (e.g. about losing money) as compared to a positive context (e.g. about winning money). We have shown that context effects also occur when context and face are presented simultaneously. From a psychological perspective this is important, because it shows that emotional information from the context and face can be combined simultaneously, and on an early stage of processing, which appears to diverge from the face recognition model by Bruce and Young (1986), in which facial expressions are extracted after a stage of structural encoding, in which context information is incorporated at a relatively late semantic stage of processing.

The N170 may (in part) be generated by the fusiform gyrus (Pizzagalli et al., 2002). The N170 amplitude to faces has been related to the BOLD response in the fusiform gyrus (Idiaka et al., 2006). The time-course of the N170, and the generation of this component in the fusiform gyrus, is consistent with a feedback modulation from the amygdala. An intracranial study has found that the amygdala may respond differentially to facial expressions at around 120 ms (Halgren et al., 1994). Anatomical connections have been found between the amygdala and the fusiform gyrus (Aggleton et al., 1980), and in patients with amygdalar sclerosis it has been shown that the amygdala is critical in enhancing the fusiform gyrus response to facial expressions (Vuilleumier et al., 2004).

The input from the amygdala to early visual areas may importantly shape emotion perception (Halgren et al., 1994;...


