Single-trial EEG–fMRI coupling of the emotional auditory early posterior negativity

Fern Jaspers-Fayer a,b, Matthias Ertl a,c, Gregor Leicht a, Anne Leupelt d, Christoph Mulert a,*

a Psychiatry Neuroimaging Branch, Department of Psychiatry and Psychotherapy, University Medical Centre of Hamburg-Eppendorf, Neuroimaging North, Hamburg, Germany
b Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada
c Institute of Biomedical Engineering, Ilmenau University of Technology, Ilmenau, Germany
d Faculty of Medicine, Ludwig-Maximilians University, Munich, Germany

A R T I C L E   I N F O

Article history:
Accepted 6 May 2012
Available online 11 May 2012

Keywords:
Simultaneous EEG–fMRI
Auditory
Emotion
Early posterior negativity
Superior parietal lobule

A B S T R A C T

Event-related potential (ERP) studies in the visual domain often report an emotion-evoked early posterior negativity (EPN). Studies in the auditory domain have recently shown a similar component. Little source localization has been done on the visual ERP, and no source localization has been done on the auditory EPN. The aim of the current study was to identify the neural generators of the auditory EPN using EEG–fMRI single-trial coupling. Data were recorded from 19 subjects who completed three auditory choice reaction tasks: (1) a control task using neutral tones; (2) a prosodic emotion task involving the categorization of syllables; and (3) a semantic emotion task involving the categorization of words. The waveforms of the emotion tasks diverged from the neutral task over parietal scalp during a very early time window (132–156 ms) and later during a more traditional EPN time window (252–392 ms). In the EEG–fMRI analyses, the variance of the voltage in the earlier time window was correlated with activity in the medial prefrontal cortex, but only in the word task. In the EEG–fMRI analyses of the traditional EPN time window both emotional tasks co-varied with activity in the left superior parietal lobule. Our results support previous parietal cortex source localization findings for the visual ERP, and suggest enhanced selective attention to emotional stimuli during the EPN time window.

© 2012 Elsevier Inc. All rights reserved.

Introduction

The brain rapidly and efficiently allocates resources to emotionally salient stimuli (Pourtois et al., 2012). The exact timing (Olofsson et al., 2008) and neural generators (Phan et al., 2002) involved in processing emotion have separately become intense areas of study, but only recently have researchers started to integrate findings from these two fields. As such, one of the main challenges in understanding the brain mechanisms involved in the processing of emotional events remains the combination of techniques that record with high temporal resolution and complementary techniques that record with high spatial resolution (Brosch et al., 2011; Junghöfer et al., 2006).

Event-related potentials (ERP), derived from electroencephalograms (EEG), offer a temporal resolution of 1 ms or better (Luck, 2005), and can give timing information about emotional events (Olofsson et al., 2008). Electrophysiological studies of brain activity have consistently demonstrated ERP differences for emotional compared to neutral visual stimuli at both early (Schupp et al., 2006) and late (Hajcak et al., 2010) stages. One of the most robust results is an early posterior negativity (EPN), which has been reported in a wide range of emotional paradigms using visual stimuli, and recently in two tasks that used auditory stimuli (Mittermeier et al., 2011). Researchers theorize that the EPN indexes automatic early selective attention to motivationally relevant emotional stimuli compared to neutral stimuli (for review see Schupp et al., 2006).

Despite consistent reports of an EPN effect, there is a considerable debate over the generators of the visual EPN (for reviews see Junghöfer et al., 2006; Olofsson et al., 2008). In part this is due to the voltage topography of the effect, which is traditionally described as a negative potential shift over occipital–temporal scalp, in response to emotional compared to neutral stimuli (Junghöfer et al., 2001), but is occasionally described as an early anterior positivity (EAP) over fronto-central sites (e.g. Sauter and Eimer, 2010; Taake et al., 2009), particularly when a linked-mastoid or linked-ear reference is used (Junghöfer et al., 2006). The use of average reference, the calculation of current source density (CSD), and mathematical source localization algorithms, such as LORETA (low-resolution tomography; Pascual-Marqui, 2002) could help us to estimate the neural generators of the EPN, but ultimately electrophysiological measures cannot provide the same spatial precision as hemodynamic measures.

Functional magnetic resonance imaging (fMRI), with spatial resolution on the order of a millimeter, has shown that emotional stimuli, compared to neutral stimuli, elicit an increase in the blood oxygen
level dependent (BOLD) signal in a number of brain structures (Phan et al., 2002), but the timing and interaction of these structures are unknown. It is possible, however, that the networks seen in fMRI may be related to ERP components measured at the scalp (Mulert and Lemieux, 2010).

Thus, it has been suggested that simultaneous EEG and fMRI data could give us a much better understanding of the underlying neural generators involved in emotion processing in the human brain (Schielen and Schäfer, 2009). To date conjoint EEG–fMRI has only once been used to study late ERP emotion components. Simultaneous EEG–fMRI has never been used to find the neural generators of the visual or auditory EPN. In fact, to our knowledge, only three attempts have been made at EPN source localization (Junghöfer, et al., 2006; Junghöfer, et al., 2010; Schupp et al., 2003). All studies estimated that bilateral occipito-temporal and occipito-parietal cortical areas generated the visual EPN.

The present experiment therefore recorded the brain’s response to emotional auditory stimuli, using simultaneously recorded EEG and fMRI. EEG and fMRI signals were combined by deriving regressors from the high-temporal resolution EEG during the EPN time window, and using those regressors to predict the BOLD signal specifically related to variations in the EPN (Mulert and Lemieux, 2010). The primary aim of the current study was to determine if there was a relationship between the scalp-recorded EPN and the related fMRI BOLD response. Based on the source localization literature we hypothesized that the EPN would be related to activity in the parietal cortex during emotional tasks.

**Methods**

Our research protocol was approved by the ethics committee for human experiments at the Ludwig-Maximilians University, Munich, Germany, and was carried out in accordance with the Declaration of Helsinki. The experimental procedures were explained to all participants, and declarations of consent were always signed before the experiment began. Participants were paid €25 for their participation.

**Subjects**

Nineteen healthy volunteers (10 women; age = 24.74, SD = 2.10; years of education = 17.31, SD = 2.66; one left-hander) took part in the simultaneous EEG–fMRI study. Subjects had no history of neurological or psychiatric disorders, and no hearing loss.

**Behavioral tasks**

Three auditory choice reaction tasks were run (see Fig. 1) in separate randomized blocks. Stimuli were presented using the BrainStim software package (Brain Products, Munich). Before recording began, a short practice run was carried out to allow the subjects to identify low (800 Hz) and high (1300 Hz) tones.

In the control task these low tones (800 Hz) and 70 high tones (1300 Hz), all 250 ms in duration, were presented in a pseudo-randomized sequence. All tones were followed by an inter-stimulus interval (ISI) of 2750, 5750 or 8750 ms. On hearing a low tone the participant pressed the left button with their left thumb, and on hearing the high tone the participant pressed the right button with their right thumb, as in previous studies run by our lab (Mulert et al., 2003; 2005; 2008).

In another block, five different syllables (ba, be, bi, bo, and bu) were presented. Seventy of these syllables had positive/happy prosody and 70 had negative/sad prosody. The validation of these syllables has been described elsewhere (Mittermeier et al., 2011), but briefly, in this stimulus set the probability of correctly recognizing a syllable with positive prosody was 99.7%, and the probability of correctly recognizing a syllable with a negative prosody was 99.6%. All syllables were 250 ms in duration, and were presented in a pseudo-randomized sequence. Again each stimulus was followed by an ISI of 2750, 5750 or 8750 ms. Participants responded to the positive by pressing a button with their right thumb, and responded to negative syllables by pressing a button with their left thumb.

Finally, five words with positive semantic meaning (happy, cheerful, friendly, calm, wonderful) and five words with negative semantic meaning (sad, irritating, terrible, disgusting, unhappy) were spoken by a monotonous female voice. Each stimulus was repeated 14 times, giving 70 positive stimuli and 70 negative stimuli in total. All stimuli lasted for 500 ms, and were presented in a pseudo-randomized sequence. The ISI was 2500, 5500 or 8500 ms long. Like in the syllable task, participants pressed the right button when they heard positive stimuli and the left button when they heard negative stimuli.

Between the tasks a short 1–2 min break was taken. Each task took approximately 15.5 min to complete. During recording participants were asked to stay calm and keep their eyes shut to reduce the number of artifacts related to eye movements in the ERP data (Winterer et al., 1997).

**Acoustic environment and sound delivery system**

Details about the acoustic environment have been described in greater detail elsewhere (Mulert et al., 2005; 2008; 2010). Binaural

---

**Fig. 1.** All three auditory choice reaction tasks followed the same basic template. Stimuli were presented during 140 of the MR volumes, the other 160 volumes recorded trials with no auditory stimulation. The tones were either high or low. The syllables were presented with either a happy or sad prosody. The duration of the word stimuli was longer than the tone or syllable stimuli (500 ms not 250 ms). All word stimuli were adjectives with either a happy or sad semantic meaning.
sound transmission was performed using an air-tube based sound delivery system (Resonance Technology, Inc., Van Mays, USA). Attenuation of the MR system noise was achieved through the use of invert earphones and circumaural earmuffs. This system attenuated some frequencies more than others. As such, before each recording, a calibration procedure was performed using an acoustic calibrator (B&K 4230). Volume was then digitally adjusted to achieve the desired volume level of 85 dB SPL for all participants before each run.

Behavioral analysis

Behavioral analyses were performed in SPSS (version 19.0 for the Macintosh) with \( \alpha = 0.05 \). One-way repeated measures ANOVAs were used to test for behavioral effects. In the first behavioral analysis Task (tones, syllables, words) was used as the within-subject factor and average reaction time (RT) was used as the dependent variable. In the second analyses, Task was used as the within-subject factor and average percentage of trials correct was used as the dependent variable. For all one-way repeated measures ANOVA analyses Greenhouse-Geisser corrections were applied to the degrees of freedom (Jennings and Wood, 1976) and corrected \( p \)-values have been reported. If an ANOVA was significant, Bonferroni corrected paired-sample t-tests were conducted to further explore the significant main effects and interactions.

**EEG — removal of ballistocardiogram artifacts**

Ballistocardiogram (BCG) artifacts are detrimental to EEG analyses (Ertl et al., 2010). Therefore, BCG artifacts were detected and marked using BrainVision Analyzer software (version 2.0, Brain Products, Munich, Germany), before data were exported to EEGLAB (version 6.01b Delorme and Makeig, 2004), an open source toolbox developed by Schwartz Center for Computational Neurosciences (La Jolla, CA; http://www.sccn.ucsd.edu/eeeglab) running on MATLAB. The FMRIB plug-in for EEGLAB was then used to remove BCG artifacts using Optimal Basis Set (OBS) using two principal components (Niazy et al., 2005). OBS has been found to efficiently remove BCG artifacts and substantially improve the quality of EEG signals recorded inside the scanner (Debener, et al., 2007; Vanderperren, et al., 2010).

**ERP — data preprocessing**

ERP data were returned to BrainVision Analyzer (version 2.0, Brain Products, Munich, Germany) for further filtering, re-referencing, segmentation, baselineing, artifact rejection and averaging. A digital high-pass filter (0.1 Hz, 12 dB/oct) was applied to the data. No low-pass filter was applied at this time, as the data were low-pass filtered earlier. Data were re-referenced to the average reference, and Cz was reused as an EEG channel. Segmentation was based on stimulus onset (from 200 to 650 ms), and the resulting epochs were aligned to a −200 to 0 ms baseline. Trials with artifacts in any encephalographic channel were detected automatically, using a ± 100 \( \mu \)V criterion and visual inspection, as suggested by Picton et al. (2000) Two participants’ EEG dataset were not useable because of residual artifacts in the recordings. Therefore, all ERP results were calculated on a sample of 17 participants.

**ERP — regions of interest on the scalp**

To define our electrode and time windows of interest we ran one-way repeated measures ANOVAs on all electrodes. Task was used as the between subject factor and the average voltage amplitude during the 100–500 ms time window was used as the dependent variable. Pz showed the highest F-value for Task, \( F(2,32) = 18.86, p = 0.00005 \), Greenhouse–Geisser corrected, compared to all other electrodes (see Fig. 2).

**FMRI — acquisition**

Imaging was performed in a 1.5 Tesla Siemens Magnetom Sonata scanner with echo-planar capability. For anatomical registration of the functional data high-resolution anatomical databases were acquired using a three-dimensional T1-weighted sequence (MPRAGE). During the functional imaging session, 12 slices were acquired covering the whole brain with a T2*-weighted EPI sequence (MPRAGE). During the functional imaging session, 12 slices were acquired covering the whole brain with a T2*-weighted EPI sequence (MPRAGE). During the functional imaging session, 12 slices were acquired covering the whole brain with a T2*-weighted EPI sequence (MPRAGE).
resulting pixel size = 2.8 × 2.8 mm) in the same position as the anatomical images. During the experiment, the stimuli were presented between slice acquisition intervals, in order to reduce the influence of the scanner noise on stimulus presentation (see Fig. 1).

The participant’s head was immobilized during the experiment using pads adjusted to the individual head size, thus minimizing involuntary head movements. Three images at the beginning of each run were discarded due to inhomogeneities of the magnetic field (T1 saturation effects), leaving a total of 300 scans for each task.

**FMRI — preprocessing**

BrainVoyager QX (version 2.3.1.1770; Brain Innovation, Maastricht, Netherlands) was used to analyze the recorded MRI data (Goebel et al., 2006). To minimize the effect of head movements on the subsequent data analysis all volumes were realigned to the first volume (tri-linear interpolation), using a reduced set of voxels (12%) in each volume, to decrease processing time. Correction for slice time errors (sinc interpolation), Gaussian 3D spatial smoothing (8 mm kernel size, full-width half-maximum) and high-pass temporal filtering (5 cycles in time course) was applied. Finally, functional data were transformed into Talairach space (Talairach and Tournoux, 1988), and aligned with the three-dimensional anatomical volumes from the same session and interpolated to a resolution of 3 × 3 × 3 mm.

Three participants’ fMRI datasets were not used because of excessive artifacts in the recordings. The fMRI analyses have therefore been calculated only on those participants who produced good-quality EEG and fMRI recordings. Hence, all fMRI results are reported on 14 participants.

**FMRI — analyses**

We used a multi-study random fixed effect general linear model (RFX-GLM), with predictor time courses for the tones, syllables and word conditions. As the tasks were completed in different runs, we effects coded the predictors so ensure that the order was the same for all subject-level design matrices before entry into the multi-study design matrix. We also included a constant term to model the baseline signal level. These three main predictors plus the constant term were then used to estimate the individual z-normalized time courses for each voxel belonging to each subject. A whole brain mask, created from a Talairach-transformed representative subject in the sample, was applied and then a 2–Gamma hemodynamic response function (HRF) was used to convolve the regressors. At the voxel level BOLD signal changes were reported at $p < 0.005$ for the standard fMRI analyses, and at $p < 0.05$ for the EEG–fMRI analyses. In all fMRI analyses multiple comparisons were corrected for using cluster-size thresholding at the $p < 0.05$ level. For each contrast we used 1000 iterations to determine the true false-positive rate. Constraining the analysis on the basis of cluster size allowed us to control for multiple comparisons without the concomitant loss of power that would occur with other correction methods, such as Bonferroni correction (Forman et al., 1995). Labeling of anatomical regions was based on the Talairach Daemon (http://www.talairach.org/daemon.html).

Creating EEG regressors for single-trial analysis

To create regressors for the EEG–fMRI analyses, each trial in each task provide two average voltage amplitudes at Pz, one for the early (132–156 ms) and one for the traditional EPN (252–392 ms) time window. Artifacts were defined as any amplitude that exceeded the mean amplitude of that time window in that task by more than 3 standard deviations. The number of artifacts according to this definition was small (<5 in each data set). When an artifact occurred in one of the trial’s time windows it was replaced by the participant’s average value during that time window for that task.

For both the early and traditional EPN time windows the time-series of voltage amplitudes were orthogonalized with respect to the target function (Schmidt-Gram orthogonalization) in order to detect hemodynamic responses specifically related to variations in the ERP response and not to some general feature of the target detection process (see Mulert et al., 2008 for further details).

**EEG–fMRI GLM**

For both emotional tasks (syllables, words) the orthogonalized time-series of the voltage amplitudes during both time windows of interest (early, traditional EPN) were used to calculate regressors for the BOLD response (Fig. 4). As with the regressors used in the
traditional fMRI analysis, the four EEG–fMRI regressors were convolved with a 2-Gamma HRF, z-normalized, and effects coded for each subject. Again, a constant term was included to model the baseline signal level. Subject-level design matrices for each subject were entered into a multi-study random fixed effect general linear model (RFX-GLM). The same mask, threshold values and correction methods that were used as in the standard fMRI analyses above.

Results

Behavioral results

The average reaction time (RT) and percentage of correct trials in each task can be found in Table 1. In the RT analysis there was a significant effect of Task, $F(2,36) = 79.98$, $p = .001$. This was caused by a significant difference between the syllable and tone conditions, $t(18) = 10.37$, $p = .001$, and a significant difference between the word and tone conditions, $t(18) = 10.62$, $p = .001$, such that it took longer to respond to trials in emotional runs than in neutral runs. There was no significant difference between the syllable and word tasks, $t(18) = 1.23$, $p = .236$. In the percentage of correct trial analysis there was a significant effect of Task, $F(2,36) = 6.69$, $p = .004$, driven primarily by the difference between the syllable and tone conditions, $t(18) = -3.29$, $p = .004$, the difference between words and tones did not survive Bonferroni correction, although it was in the same direction.

ERP results

After running repeated measures one-way ANOVAs to define the electrode (Pz) and windows of interest (above) we ran post hoc paired $t$-tests to further explore our results. These analyses revealed that during the first time window (132–156 ms) the syllable task (average = $-1.00 \mu V$, SD = $1.32 \mu V$; $t(16) = -4.44$, $p = .001$) and word task (average = $-1.02 \mu V$, SD = $1.25 \mu V$; $t(16) = -4.35$, $p = .001$), elicited much larger negativities than the tone task (average = $0.49 \mu V$, SD = $1.78 \mu V$). There was no significant difference between syllables and words $t(16) = -0.40$, $p = .969$.

Similarly, during the second time window (252–392 ms), the syllable (average = $3.79 \mu V$, SD = $2.98 \mu V$; $t(16) = -3.40$, $p = .004$) and word tasks (average = $1.63 \mu V$, SD = $2.89 \mu V$; $t(16) = -9.25$, $p = .001$), elicited waveforms that were more negative than the tone task (average = $5.01 \mu V$, SD = $3.09 \mu V$). Additionally the difference between syllables and words in this instance was significant, $t(16) = -6.32$, $p = .001$, in that words producing more negative voltages than syllables.

Table 2 shows the results for each task compared to the baseline signal modeled by the constant. To summarize the standard fMRI results, all tasks showed significant bilateral activations of the superior temporal gyri (STG), superior temporal sulci (STS), dorsal anterior cingulate cortex (ACC) and supplementary motor area (SMA), as anticipated for these auditory choice reaction tasks (Mulert et al., 2008).

Table 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>RT (SD)</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tones</td>
<td>491.08 (123.46)</td>
<td>97.97%</td>
</tr>
<tr>
<td>Syllables</td>
<td>723.50 (175.02)</td>
<td>93.23%</td>
</tr>
<tr>
<td>Words</td>
<td>748.74 (203.53)</td>
<td>96.95%</td>
</tr>
</tbody>
</table>
Table 2

Results for the Random Effect Analysis (RFX) multi-study GLM. Each active task is contrasted against the baseline signal (as modeled by the constant). All contrasts were thresholded for the standard fMRI analyses at $p<0.05$ at the voxel level and $p<0.05$ at the cluster level. L = left; R = right; B = bilateral; M = midline; F = frontal; temp = temporal; sup = superior; inf = inferior; mid = middle; med = medial; post = posterior; ACC = anterior cingulate cortex; PCC = posterior cingulate cortex; SMA = supplementary motor area; IFG = Inferior frontal gyrus.

Table 3

This table displays the results for the single-trial coupling EEG–fMRI analyses. Contrasts were thresholded at $p<.05$ at both the voxel level and the cluster level. L = left; R = right; B = bilateral; ant = anterior; front = frontal; temp = temporal; sup = superior; mid = middle; PCC = posterior cingulate cortex; OFC = orbital frontal cortex.

Single-trial coupling results

Again, predictors were compared against the baseline signal modeled by the constant. To summarize the EEC–fMRI results, the variability of the EEG response in the first time window was not significantly related to the BOLD activity in the syllable task. In the word task, however, we found bilateral activation of the orbital frontal cortex (BA 11) and the anterior prefrontal cortex (BA 10; see Table 3 and Fig. 5).

In the more traditional EPN time window (252–392 ms), the variability in the EEG for both the syllable and word task was related to BOLD activity in the left superior parietal lobule (BA 7; See Table 3 and Fig. 6).

Discussion

An early posterior negativity (EPN) for emotional compared to neutral stimuli is consistently reported in visual ERP studies, and has recently been reported in an auditory ERP study of emotion (Mittermeier, et al., 2011). Despite the fact that the EPN is a reliable finding in the field of affective psychophysiology, attempts to source localize the EPN in the visual domain are infrequent, and source localization has never been done in the auditory domain. We investigated the neural correlates of the auditory EPN by recording simultaneous EEG and fMRI activation.
Our secondary finding was a very early effect (132 ms–156 ms), related to BOLD activity in the ventromedial prefrontal cortex during the word task. Prominent theories of emotion processing in the auditory domain suggest that prosodic (Wildgruber et al., 2006) and possibly semantic (Schirmer and Kotz, 2006) emotion is processed during three successive stages. First primary and higher order acoustic regions extract information, then that information is represented in the auditory “what” pathway (particularly the right posterior superior temporal sulcus), before being evaluated in frontal cortex (Brück et al., 2011). Similar to other recent findings (Paulmann et al., 2010), our results support this model of prosodic processing, but suggest a more complicated interaction of anterior and posterior structures during the semantic processing of emotional stimuli. In the case of semantic emotion information the orbital frontal cortex (OFC) seems to be involved at very early stages.

The activation of the OFC during semantic emotion processing is similar to studies using emotional visual stimuli, where the OFC and amygdala are often found to contribute to early processing stages, long before explicit stimulus categorization. For instance, neuroanatomical single cell recordings have shown that the neurons in the ventral prefrontal cortex of healthy adults respond 120–160 ms post stimulus onset (Kawasaki et al., 2001; Pourtois et al., 2010) to emotional stimuli. Additionally, in response to visual stimuli, Carretié et al. (2006), using LORETA, were able to estimate that the ventromedial prefrontal cortex (BA 10) was the generator of a similar ERP component. We hypothesize that the OFC is activated very early during the processing of semantic emotion, triggering re-entrant signals that influence the ongoing perceptual processing of stimuli in the standard sensory pathways, similar to what occurs in the visual domain (Pourtois et al., 2012). A limitation of this study is that it was not optimized for finding these early frontal effects. Further EEG–fMRI studies should be conducted to determine the exact task and stimulus characteristics that elicit this effect, in both visual and auditory paradigms.

In conclusion, this is the first simultaneous EEG–fMRI coupling study to look at early processing of emotional stimuli in the auditory domain, capitalizing on the high temporal resolution of ERPs and the high spatial resolution of fMRI. We found an EPN in both prosodic and semantic emotion tasks, and found a common neural generator in the superior parietal lobule. Additionally, we found that a very early component related to activity in the ventromedial prefrontal cortex during the semantic processing of emotional words. This suggests that, similar to the processing of emotional pictures, there is an early top-down influence on the processing of emotional words from prefrontal cortex. Although this was an unexpected result, we believe it should influence models of emotion processing in the auditory domain.

Acknowledgments

This study was supported by the Faculty of Medicine, Ludwig-Maximilians-University, Munich, Germany (Foerderprogramm fuer Forschung und Lehre; 572) and grants from the Canadian Institutes of Health Research (FJF) and the Michael Smith Foundation for Health Research (FFJ). The authors would like to thank Armin Heinze (BrainVoyager), José Raúl Naranjo Muradás, and Nikolay Novitskiy (BrainVision Analyzer) for correspondence regarding the statistical tests underlying some of our more complex analyses. The authors reported no biomedical financial interests and no potential conflicts of interest.

References
