NeuroImage 60 (2012) 553–561

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Crossmodal interactions in audiovisual emotion processing

Veronika I. Müller [a,b,c,*], Edna C. Cieslik [a,b,c], Bruce I. Turetsky [d], Simon B. Eickhoff [a,b,c,e]

[a] Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Germany
[b] Department of Neuroscience und Medicine, INM-2, Research Centre Jülich, Germany
[c] JARA — Translational Brain Medicine, Jülich/Aachen, Germany
[d] Neuropsychiatry Division, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, USA
[e] Institute of Clinical Neuroscience and Medical Psychology, Heinrich Heine University, Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 13 September 2011
Revised 16 November 2011
Accepted 3 December 2011
Available online 13 December 2011

Keywords:
fMRI
DCM
Incongruence
Amygdala
pSTS
Audiovisual

ABSTRACT

Emotion in daily life is often expressed in a multimodal fashion. Consequently emotional information from one modality can influence processing in another. In a previous fMRI study we assessed the neural correlates of audio–visual integration and found that activity in the left amygdala is significantly attenuated when a neutral stimulus is paired with an emotional one compared to conditions where emotional stimuli were present in both channels. Here we used dynamic causal modelling to investigate the effective connectivity in the neuronal network underlying this emotion presence congruence effect. Our results provided strong evidence in favor of a model family, differing only in the interhemispheric interactions. All winning models share a connection from the bilateral fusiform gyrus (FFG) into the left amygdala and a non-linear modulatory influence of bilateral posterior superior temporal sulcus (pSTS) on these connections. This result indicates that the pSTS not only integrates multi-modal information from visual and auditory regions (as reflected in our model by significant feed-forward connections) but also gates the influence of the sensory information on the left amygdala, leading to attenuation of amygdala activity when a neutral stimulus is integrated. Moreover, we found a significant lateralization of the FFG due to stronger driving input by the stimuli (faces) into the right hemisphere, whereas such lateralization was not present for sound-driven input into the superior temporal gyrus.
In summary, our data provides further evidence for a rightward lateralization of the FFG and in particular for a key role of the pSTS in the integration and gating of audio–visual emotional information.

Introduction

The ability to accurately extract emotional information from facial and vocal expressions is an important part of social functioning. In everyday life, however, we are rarely confronted with information from only one sensory channel. Rather, emotional processing is most of the time based on multimodal inputs, for instance seeing a happy face and concurrently hearing laughter. Consequently, processed unimodal inputs from different channels have to be integrated into a coherent construct of emotional perception.

Studies of audiovisual integration have predominantly assessed the neural correlates of audiovisual speech perception (Beauchamp et al., 2010; Benoit et al., 2010; Calvert et al., 2000; Sekiyama et al., 2003) as well as object and action processing (Beauchamp et al., 2004; James et al., 2011; Taylor et al., 2006). Yet, some authors have also investigated the integration of emotional information (Hagan et al., 2009; Kreifelts et al., 2007, 2009; Pourtois et al., 2005; Robins et al., 2009). Independent of the different stimuli and experimental designs, all of these studies identified the posterior temporal sulcus (pSTS)/middle temporal gyrus (MTG) as a key region for audiovisual integration. Although some authors could not replicate this finding (Hocking and Price, 2008; Olson et al., 2002), other studies assessing effective connectivity (Kreifelts et al., 2007) and using TMS (Beauchamp et al., 2010) further support the role of the pSTS in the processing and integration of multimodal stimuli. For instance, Kreifelts et al. (2007) report increased effective connectivity from FFG and STG to ipsilateral pSTS in multimodal compared to unimodal conditions.

Unimodal emotional processing in turn has been extensively studied, with a high proportion of studies reporting amygdala activation for both visual and auditory emotion processing (Belin et al., 2004; Fecteau et al., 2007; Gur et al., 2002; Habel et al., 2007; Phillips et al., 1998; Sabatinelli et al., 2011; Sander and Scheich, 2001). Crossmodally Dolan et al. (2001) and Klasen et al. (2011) used an emotional audiovisual paradigm and found greater activity in the amygdala in emotional congruent compared to incongruent conditions. In a recent fMRI study investigating audiovisual emotional integration we differentiated between two different types of incongruence processing, incongruence of emotional valence and incongruence of emotion presence (Müller

* Corresponding author at: Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Pauwelsstraße 30, D-52074 Aachen, Germany. Fax: +49 241 80 82401.
E-mail address: vmueller@ukaachen.de (V.I. Müller).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.12.007

et al., 2011). While incongruence of emotional valence refers to pairing of two emotional stimuli which differ in valence (e.g., positive sound such as laughter and a negative visual stimulus such as a fearful face), incongruence of emotional presence is defined as presentation of an emotional stimulus in one channel and a neutral one in the other (e.g., an emotional sound such as laughter or scream and a neutral visual stimulus such as a face showing no emotional expression). While no effect in the amygdala could be observed for (in)congruence of emotion valence, we found a statistically significant effect for the congruence of emotional presence. In particular, activity in the left amygdala was significantly higher in conditions where an emotional stimulus was presented in both channels (emotion presence congruence) compared to those conditions where an emotional stimulus in one modality was paired with a neutral stimulus in the other channel (Müller et al., 2011). This result indicates that independent of crossmodal *valence* congruency, left amygdala activity is elevated when an emotional signal is not paired with neutral information. This emotion presence congruence (EPC) effect may be understood as differential additive and competitive effects on amygdala activity in audiovisual integration. Emotional stimuli in both channels, congruent or incongruent in valence, both cause amygdala activation resulting in summation of activity. In contrast, an emotional stimulus paired with a neutral one leads to competition and the neutral information in one channel may counteract activation by the emotional stimulus in the other modality. This result indicates that the amygdala not only signals the presence of emotional, but also the absence of neutral information, suggesting a more general role of the amygdala in the evaluation of social relevance (Adolphs, 2010). However, the neural interactions of the network leading to a mechanistic understanding of this signal in the amygdala can't be uncovered by conventional fMRI analysis. Using Dynamic Causal Modelling (DCM), the present study therefore aims to investigate the effective connectivity between uni- and multimodal sensory regions contributing to the modulation of left amygdala activity by emotion presence congruence (EPC). In total 48 different models, reflecting all different neurobiological hypotheses, were varied along three different factors: i) presence and absence of interhemispheric connections, ii) region projecting into the left amygdala and iii) modulation of the effective connectivity towards the left amygdala. Seven regions were included into the models, bilateral fusiform (FFG) and superior temporal gyri (STG) as face and voice sensitive areas (Belin et al., 2000; Gainotti, 2011; Grill-Spector et al., 2004; Kanwisher et al., 1997; Kanwisher and Yovel, 2006) respectively as well as bilateral pSTS because of its role in audiovisual processing (Beauchamp et al., 2004, 2010; Calvert et al., 2000; Hagan et al., 2009; James et al., 2011; Kreifelts et al., 2007, 2009). Finally the left amygdala was included as the aim of the present study is to investigate the neural interactions which lead to the EPC effect found in that region.

Methods

Subjects

The present effective connectivity modeling was based on a fMRI experiment investigating audiovisual emotion integration (Müller et al., 2011). Six participants of the original sample of 35 subjects were excluded from the DCM analysis due to missing activation on the single-subject level in one or more regions assessed by the dynamic causal models, leaving 29 subjects for the DCM. Two subjects were excluded after the modeling, as their model parameters were strongly deviant (more than two standard deviations from the group mean, leading to violations of the normality assumption). In order to avoid any bias by these subjects, all group analyses were based on the remaining 27 subjects. All participants had normal or corrected-to-normal vision and were right handed as confirmed by the Edinburgh Inventory (Oldfield, 1971). No subject had a history of any neurological or psychiatric disorder, including substance abuse. Informed consent for the study, which was approved by the ethics committee of the School of Medicine of the RWTH Aachen University, was given by every participant.

Stimuli and procedure

A detailed description of the stimulus material is reported in Müller et al. (2011). In brief, we used a total of 10 (5 male and 5 female) different faces, each showing three different expressions (happy, neutral, fear). In addition, 10 different neutral faces blurred with a mosaic filter were included as masks. The auditory stimuli were neutral (yawning) or emotional sounds like laughing (happy) or screaming (fear), with 10 sounds for each category (5 male and 5 female in each case).

In total 180 audiovisual stimulus pairs were presented. Every trial started with the presentation of a sound concurrently with a blurred neutral face. After 1000 ms the blurred face was displaced by the target stimulus (neutral or emotional face) and this was presented together with the ongoing sound for another 500 ms. The subjects were asked to ignore the sound and to rate the presented face on an eight-point rating scale from extremely fearful to extremely happy. Each pair was hence presented for 1500 ms. Between trials a blank screen was presented for 4000–6000 ms (uniformly jittered) before the onset of the next trial. All stimuli were presented with the software Presentation 14.2 (http://www.neurobs.com/) and responses were given manually using an MR-compatible response pad.

fMRI data acquisition and pre-processing

Images were acquired on a Siemens Trio 3T whole-body scanner (Erlangen, Germany) in the Research Center Jülich using blood-oxygen-level-dependent (BOLD) contrast [Gradient-echo EPI pulse sequence, TR = 2.2 s, in plane resolution = 3.1 × 3.1 mm, 36 axial slices (3.1 mm thickness)] covering the entire brain. To allow magnetic field saturation, prior to image acquisition 4 dummy images were collected which were discarded before further processing. Images were analyzed using SPM5 (www.fil.ion.ucl.ac.uk/spm). The data was pre-processed by using head movement correction, spatial normalization to the MNI single subject space, resampling at $2 \times 2 \times 2$ mm^3 voxel size and smoothing using an 8 mm FWHM Gaussian kernel (for a detailed description see Müller et al., 2011).

Statistical analysis (GLM)

The imaging data were analyzed using a General Linear Model. Each experimental condition ("fearful face + scream", "fearful face + yawn", "fearful face + laugh", "neutral face + scream", "neutral face + yawn", "neutral face + laugh", "happy face + scream", "happy face + yawn", "happy face + laugh") and the (manual) response were separately modeled by a boxcar reference vector convolved with a canonical hemodynamic response function and its first-order temporal derivative. Simple main effects for each of the nine experimental conditions and the response were computed for every subject and then fed to a second-level group-analysis using an ANOVA (factor: condition, blocking factor subject) employing a random-effects model. By applying appropriate linear contrasts to the ANOVA parameter estimates, simple main effects for each condition (versus baseline) as well as comparisons between the conditions were tested. The resulting SPM (T) maps were then thresholded at $p<.05$, corrected for multiple comparisons by controlling the family-wise error (FWE) rate according to the theory of Gaussian random fields (Worsley et al., 1996). Ensuing activations were anatomically localized using version 1.6b of the SPM Anatomy toolbox (Eickhoff et al., 2005, 2006, 2007; www.fz-juelich.de/ime/spm_anatomy_toolbox).