NeuroImage 61 (2012) 1255–1267

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Disentangling the brain networks supporting affective speech comprehension

Pierre-Yves Hervé [a,b,c,*], Annick Razafimandimby [d], Mathieu Vigneau [e], Bernard Mazoyer [a,b,c,f], Nathalie Tzourio-Mazoyer [a,b,c]

[a] Univ. Bordeaux, Groupe d'Imagerie Neurofonctionnelle, UMR 5296, F-33000 Bordeaux, France
[b] CNRS, Groupe d'Imagerie Neurofonctionnelle, UMR 5296, F-33000 Bordeaux, France
[c] CEA, Groupe d'Imagerie Neurofonctionnelle, UMR 5296, F-33000 Bordeaux, France
[d] ISTCT, UMR 6301 CNRS, CEA, Université de Caen, France, GIP CYCERON, Blvd Becquerel, BP 5229,14074 Caen Cedex France
[e] FRE 3481 CNRS, Université de Reims Champagne-Ardennes, équipe MéDIAN, Faculté de Médecine de Reims, Reims, France
[f] Institut Universitaire de France, Paris, France and CHU Caen Côte de Nacre, France

ARTICLE INFO

Article history:
Accepted 18 March 2012
Available online 30 March 2012

Keywords:
fMRI
Theory of mind
Emotion
Language
Functional connectivity
Replicator dynamics

ABSTRACT

Areas involved in social cognition, such as the medial prefrontal cortex (mPFC) and the left temporo-parietal junction (TPJ) appear to be active during the classification of sentences according to emotional criteria (happy, angry or sad, [Beaucousin et al., 2007]). These two regions are frequently co-activated in studies about theory of mind (ToM). To confirm that these regions constitute a coherent network during affective speech comprehension, new event-related functional magnetic resonance imaging data were acquired, using the emotional and grammatical-person sentence classification tasks on a larger sample of 51 participants. The comparison of the emotional and grammatical tasks confirmed the previous findings. Functional connectivity analyses established a clear demarcation between a "Medial" network, including the mPFC and TPJ regions, and a bilateral "Language" network, which gathered inferior frontal and temporal areas. These findings suggest that emotional speech comprehension results from interactions between language, ToM and emotion processing networks. The language network, active during both tasks, would be involved in the extraction of lexical and prosodic emotional cues, while the medial network, active only during the emotional task, would drive the making of inferences about the sentences' emotional content, based on their meanings. The left and right *amygdalae* displayed a stronger response during the emotional condition, but were seldom correlated with the other regions, and thus formed a third entity. Finally, distinct regions belonging to the Language and Medial networks were found in the left angular gyrus, where these two systems could interface.

Introduction

Speech can convey emotions through two different channels: the voice, with prosody (the melody of speech), and the verbal content, including both the particular choice of words (lexical content) and the overall meaning of the sentence, which emerges from the integration of lexical, grammatical and contextual information. In daily language use, these pragmatic and linguistic aspects of affective discourse have to be combined by the receiver for him to build an accurate representation of the sender's emotional state.

Cerebral imaging has shown that whether one is attending to or producing speech, audio-visual speech, sign-language, or written language, and depending on the semantic categories involved (Martin and Chao, 2001), the performance of a language task will rely on interactions between the same network of core language-related areas and several other networks of specialized auditory, visual, spatial or emotional areas (Desai et al., 2010; MacSweeney et al., 2008). Beyond the classical perisylvian language areas, an involvement of dorsal and ventral medial areas of the prefrontal cortex (mPFC), the posterior cingulate cortex (pCC), the posterior STS and the anterior temporal lobe (aTL) is observed along with the left middle temporal and inferior frontal gyrus regions during coherent text comprehension (Ferstl et al., 2008). The elaboration of text coherence corresponds to the integration of the information carried by a series of semantically related sentences, which are put together so as to form a situation model. Situation models can be defined as "mental representations of the state of affairs described in a text rather than of the text itself" (Zwaan, 1999). Such comprehension processes requires making and selecting relevant inferences. For instance, Kintsch (1988) points that upon hearing the phrase "the hikers saw the bear", listeners reach the relevant conclusion that the hikers were scared, rather than the irrelevant conclusion that they had their eyes open. This selection of the relevant inferences over the irrelevant ones often relies on prior knowledge, not present in the text. Besides, combinations of the mPFC, pCC, temporal poles and temporo-parietal junction (TPJ) areas are recruited during tasks, language-based or not (Castelli et al., 2000; Gallagher et al., 2000), which involve making inferences regarding the mental states of others (Fletcher et al., 1995). This process is known as mentalizing or theory of

* Corresponding author at: Groupe d'Imagerie Neurofonctionelle, Université Bordeaux Segalen, 146 rue Léo-Saignat, Case 71, 33076 Bordeaux Cedex, France.
E-mail address: pierre-yves.herve@u-bordeaux2.fr (P.-Y. Hervé).

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2012.03.073

mind (ToM), and is most strongly associated with the mPFC and TPJ regions (Frith and Frith, 2006; Saxe, 2006). Furthermore, using multi-voxel pattern analysis and an emotional-intensity judgment task on vocal, facial or body expressions, mPFC and left posterior STS regions were recently shown to harbor modality-independent representations of different categories of emotions perceived in others (Peelen et al., 2010).

The involvement of these two regions in the comprehension of affective speech has already been demonstrated by an fMRI study at 1.5 T on the neural bases of emotional speech comprehension (Beaucousin et al., 2007), in which the brain activations recorded during an affective sentence-classification task (is the sentence, as said by the speaker, angry, happy or sad?) were compared with those observed during a grammatical sentence-classification task (is the sentence in the 1st, 2nd or 3rd grammatical person?). Additionally, in both tasks, the presence of emotional prosody was manipulated, through the use of either an actor voice or a text-to-speech software voice (which included grammatical prosody). The presence of emotional prosody during the affective sentence classification task was associated with higher activity in voice-sensitive areas of the anterior and posterior right superior temporal sulcus (STS) and bilateral *amygdalae*. Interestingly, irrespective of prosody, the left and right inferior frontal gyrus (IFG, with a leftward asymmetry), left TPJ and left dorsal mPFC were significantly more active during the emotional than during the grammatical classification task.

The authors proposed that the activation of these regions reflected the involvement of different neural systems during the classification of emotional sentences, with an emotional prosody system including the right temporal areas and the amygdala, and a linguistic system, with the IFG being involved in the lexico-syntactic processing of emotional cues. In addition, the activation of medial and angular gyrus areas suggested the implication of a ToM system, what was supported by an *ad-hoc* meta-analysis of ToM studies. A key issue concerned the role of the left angular gyrus, which could be associated either with semantic (Seghier et al., 2010) or ToM processes (Saxe and Kanwisher, 2003). Neuropsychological experiments using false-belief tasks showed that the left TPJ was necessary for inferring the beliefs of other persons (Samson et al., 2004), and such inferential processes may be involved during the EMO task.

The purpose of the present study was to test the hypothesis that several different neural systems, related to emotional prosody, language processing or ToM, contribute to emotional speech comprehension. Accordingly, we applied a functional connectivity analysis to a new 3 T-MRI dataset acquired with event-related versions of the two tasks used in the previous report, using only the actor voice. In addition, the participants answered a debriefing questionnaire including items focusing on the use of mentalizing strategies to solve the emotional speech task.

Material and methods

Participants

We included a total of 51 participants (29 males), including 3 left-handers (2 females). The median age of the group was 28 years (range: 18–53, mean ± sd: 30.6 ± 8.1 years). The average Edinburgh Handedness Inventory score (Oldfield, 1971) was 92.9 ± 13.5 for the self-reported right-handers, whereas all 3 left-handers were at −100 on the Edinburgh scale. The average level of education was 16 ± 3.4 years (range: 11–20 years). In France, the 13th year of education is the first at university-level. We did not detect any abnormality in the structural scans of any of the included participants. The local ethics board (CCPRB: Comité Consultatif de Protection des Personnes se prêtant à la Recherche Biomédicale, Basse-Normandie) had approved the experimental protocol. All participants gave their informed, written consent, and received an allowance for their participation.

Tasks

This study relied on the experimental paradigm developed by Beaucousin et al. (2007), and included two auditory sentence-classification tasks.

In the first task, referred to as EMO for “emotional task”, the participants were asked to classify the sentences into three different categories on the basis of their emotional content (Anger, Sadness, or Happiness). In the EMO task, the sentences contained both lexical and prosodic emotional cues.

In the second task, referred to as GRAM for “grammatical task”, sentences were classified according to a grammatical feature, the grammatical person of the sentence (1st person, 2nd person, or 3rd person). The sentences used in the GRAM task were devoid of emotional information, either at the prosodic or lexical level. This grammatical task was an appropriate reference for the EMO task because it tapped sentence-level language processes, yet without triggering emotional processes, as no such content was present in the sentences. Furthermore, so as to minimize the linguistic differences between the EMO and GRAM sentences, they were matched in terms of their number of words, imageability and syntactic structures. The sentences were all composed of 2-to-3-syllables-words from the BRULEX database, which were frequent and highly conceivable. Last, the GRAM reference task also required the participants to perform a 3-choice task.

Experimental paradigm

Each participant performed two different runs of each task. Each run contained 24 sentences, and followed a slow event-related design, with the different sentence categories occurring randomly, but in the same order for all participants. Sentence duration ranged from 2 to 4.5 s (mean ± sd, EMO: 2.65 ± 0.49 s, GRAM: 2.64 ± 0.49 s). After the end of the sentence, the participants had to respond with a 4-button response pad within 3 s (against 1 s in the previous block-design study, Beaucousin et al., 2007). In order to prevent the participants from replaying the task during the inter-trial interval, after each sentence classification trial, the participants performed a “beep detection task”: they heard two different pure tones in random order, separated by 2 to 8 s, and had to respond upon hearing the lower-frequency tone. The total event duration (sentence classification plus beep detection) was 14 ± 2 s (*i.e.* 7 ± 1 TR).

The presentation of the stimuli and recording of responses were done using the E-Prime 1.2 software. The auditory stimuli were delivered via MR compatible headphones (MR-CONFON gmbh), and the manual responses were collected using an MR-compatible button box (FORP, Current Designs).

Debriefing

Shortly after the scanning session, the participants were submitted to a standardized debriefing interview. The participants had to report on their strategy during the EMO task. The experimenter recorded on a form whether or not the participant relied on lexical cues (i.e. emotional words), or on the emotional prosody (intonation). The use of mentalizing strategies was documented with three questions. We asked the participants if they took the perspective of an addressee, and/or if they projected themselves in the place of the speaker, and if they had used their own knowledge of social (interpersonal) relationships to solve the task. In order to evaluate the encoding of the sentences, and to assess the presence of an advantage in favor of the emotional material, the participants had to recognize 12 sentences that they had heard in the scanner, among a total of 24 written sentences.