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 © 2012 Elsevier Inc. All rights reserved.
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.

*Image acquisition*

Images were acquired with a Philips Achieva 3 T scanner. The protocol began with a $T_1$-weighted structural scan (3D $T_1$-TFE sequence, 256×256 matrix size with 180 slices, 1 mm isotropic resolution, sagittal slice orientation, FA = 10°, TE = 4.6 ms, TR = 20 ms, TI = 800 ms, SENSE factor = 2). A $T_2$-weighted/proton density (PD) scan was also acquired for cross-modal registration purposes (T2-FFE sequence, 256×256 matrix size with 70 slices, 2 mm isotropic resolution, axial slice orientation, TE = 35 ms, TR = 3500 ms, FA = 90°, SENSE factor = 2). Regarding the 4 fMRI runs, the EPI-BOLD sequence acquisition parameters were: 64×64 matrix with 31 slices, 3.75 mm isotropic resolution, sagittal slice orientation, TE = 35 ms, TR = 2 s, FA = 80°, no parallel imaging, and a number of 168 images for a total duration of 336 s (not including 8 dummy scans). The four different runs were ordered at random.

*Image processing*

Image analysis was performed using the SPM5 software. The $T_1$-weighted scans of the participants were normalized to a site-specific template (T-80TVS) matching the MNI space, using the SPM5 "segment" procedure with otherwise default parameters. So as to correct for subject motion during the fMRI runs, within each run, the 168 EPI-BOLD scans were realigned using a rigid-body registration. The EPI-BOLD scans then were registered rigidly to the structural $T_2$-weighted image, which was itself registered to the $T_1$-weighted scan. The combination of all registration matrices allowed warping the EPI-BOLD functional scans to the standard space with a trilinear interpolation. Once in the standard space, a 6-mm FWHM Gaussian filter was applied.

*Behavioral data analysis*

We compared the accuracy (mean number of correct responses per run, CR), the response times (RT) of correct answers, and the mean number of recognized sentences per run, between EMO and GRAM using Wilcoxon rank tests (JMP 9 software).

We computed the correlation between EMO and GRAM for RT (Pearson correlation).

After ensuring that the percentage of mean number of recognized sentences was above the chance level we compared this percentage for GRAM and EMO sentences.

We computed descriptive statistics with the participants' answers to the debriefing questions regarding the strategies they used.

*Statistical parametric mapping*

Regarding the functional imaging data, for the subject-level analyses, we used the SPM General Linear Model procedure. We specified a Finite Impulse Response (FIR) design matrix, which averaged the BOLD values across runs and events, at different intervals after the stimulus onset time (1, 3, 5, 7, 9 and 11 s post-stimulus times). The high-pass filter was set at a 128 s period.

For the group-level analyses, we used a simple random-effects model including a single factor with 12 levels (6 post-stimulus times by 2 tasks), which enabled us to compare the average response at each post-stimulus time between tasks. We first produced maps of the time course of the average BOLD response to EMO events, using one-tailed one-sample t-tests for each of the 6 time-points ($p<0.05$, family-wise error correction for multiple comparisons, FWE). Furthermore, with this model, looking at the interactions between post-stimulus times and task is a way to investigate the differences in the dynamics of the event-related responses between the two tasks. The difference in activation between t = 7 s and t = 1 s was compared between EMO and GRAM, corresponding to a time by task interaction. In the SPM group-level model it corresponded to a one-tailed interaction contrast ([EMO t = 7 minus t = 1] minus [GRAM t = 7 minus t = 1]).

*Functional connectivity*

*Regions-of-interest definition*

The first step of the functional connectivity analysis was to extract time-series at relevant locations in the brain, through a region-of-interest (ROI) approach based on in-house Matlab software. To this end, we used 4 mm-radius spheres centred on each 8-mm local maximum ("activation peak", with a 6-neighbor definition of connectivity) in the t-map of the EMO–GRAM contrast, at 7 s compared with 1 s post-stimulus time ($p = 0.05$ FWE, minimum cluster size: 5 voxels). This yielded a total of 37 ROIs, of 33 voxels each.

*Extraction of time series*

In each of these stereotaxic ROIs, we extracted the fMRI time-series (1st eigenvariate, SPM) that were recorded during the 2 EMO runs in each subject, using matlab and SPM software routines. We applied a SPM high-pass filter of 32 s period. We then removed the effects of subject motion as estimated with a linear model including the SPM canonical-HRF-convolved stimulus function and the 6 rigid-body motion parameters (high-pass filtered).

*Inter-regional correlation matrix*

We then computed the correlation matrices between the extracted time-series at the 37 ROIs (Spearman rank correlation) for each run and subject (Lohmann and Bohn, 2002). To do so, we applied Fisher's $r$ to $z$ transform to each 37 by 37 correlation matrix and then averaged the correlation coefficients across the 51 subjects and 2 replications of the EMO task. The average correlation coefficients are not distributed normally, and directly using their mean as an indicator of the central tendency of the correlation values is therefore problematic. To circumvent this issue, we applied the Fisher's $r$ to $z$ transform to all correlation coefficients before averaging them. Since these $z$-transformed values are normally distributed, their mean provides a more reliable average correlation value. The inverse transform ($z$ to $r$) was then applied to the average $z$-transformed correlation matrix, so as to obtain a more convenient matrix of $r$-values, properly averaged across subjects and replications (Lohmann and Bohn, 2002). This, as well as all of the ensuing statistical analyses, was done using the R statistical software. All three-dimensional displays of the results were performed with BrainVisa/Anatomist (Mangin et al., 2004) and/or POVray 3.6.

*Network decomposition with replicator dynamics*

We applied a replicator dynamical analysis (as described in Lohmann and Bohn, 2002) to the average correlation matrices. This iterative process allows the isolation of networks of tightly inter-correlated ROIs. It simulates the growth of a number of self-replicating entities ("replicators", here, the 37 nodes) whose fitness depends on the presence of strong interactions (here, functional correlations) with other replicators. The equation describing this "replicator" process in discrete time is as follows:

$$x_i(t+1) = x_i(t)\frac{(Wx(t))_i}{x(t)^T Wx(t)}$$

in which $W$ is a similarity matrix (symmetric and non-negative, here, our matrix of average $r$ values), with diagonal set to 0 so as to eliminate self-interaction, and $x$ is a vector containing the proportions of the $i$ replicators in the population at the $t$th iteration of the process. After convergence, the peaks forming the dominant network are removed from the correlation matrix, and the whole process is re-iterated using the smaller correlation matrix. The members of the dominant network

are defined as those with a frequency exceeding chance level, *i.e.* 1/n, with *n* being the number of replicators (Lohmann and Bohn, 2002).

After a 1st complete application of the replicator process to the matrix *W*, the correlations of the ROIs belonging to each network can be averaged, what produces a second similarity matrix, containing similarities between the networks obtained following the 1st level analysis. To explore the relationships between these networks, we ran a 2nd level replicator process on this inter-network correlation matrix (Lohmann and Bohn, 2002).

*Multidimensional scaling display of the networks*

Lohmann and Bohn (2002) used multi-dimensional scaling (MDS) as a means to "visualize the similarity structure" of the inter-regional correlation matrix *W*, and to evaluate the validity of the result of the replicator process. In the present neuroimaging context (see Friston et al., 1996; Lohmann and Bohn, 2002; Salvador et al., 2005 for applications of MDS in this domain), MDS optimizes the relative positions of the 37 points in an *ad-hoc* low-dimensional space, so that the Euclidean distance between strongly correlated ROIs is small. Therefore, in order to display the positions of the 37 regions in a "functional space", and to evaluate their consistency with the results of the replicator dynamics analysis, we computed a metric multidimensional scaling (with 3 dimensions), from a distance matrix simply defined as $d = 1 - W$. For further cross-validation, we also performed a hierarchical clustering analysis of the functional connectivity data, using the same distance matrix (results in Supplementary Fig. 4).

*Within-network partial correlation analyses*

The correlation structure of the regional data can be further described by extracting the specific connections between pairs of regions within the maximally coherent networks obtained via the replicator dynamics analysis. To do so, we computed the partial correlation matrices between the different activation foci of each network. These partial correlation matrices were derived from the corresponding average correlation matrices (using the R *corpcor* package, http://strimmerlab.org/software/corpcor). In practice, this analysis shows the relationships between pairs of regions that cannot be explained by a similar influence of any other member of the network on both these regions. Since the replicator dynamical analysis tends to produce groups of tightly correlated regions, for display purposes, we applied a descriptive arbitrary threshold of $r = 0.175$ to these partial correlation matrices.

*Regional event-related responses*

To describe the behavior of the various regions and networks during the EMO and GRAM tasks, we calculated the average event-related hemodynamic responses in each ROI using a similar FIR model as with the whole brain analyses, on the motion-corrected time-series which served for the average correlation matrices. The average responses in each ROI were sorted according to the network decomposition for the EMO task.

## Results

*Behavioral data*

For both tasks, accuracy was very high, without significant difference between EMO and GRAM (mean number of CR during EMO: 23.4 out of 24, range: 17.5–24; mean number of CR during GRAM: 23.4, range: 19.5–24; mean EMO–GRAM difference: −0.05, standard deviation: 0.73; Wilcoxon test: $p = 0.55$).

The RT of correct responses was significantly higher during EMO as compared with GRAM (mean RT during EMO: 714.5 ms, standard deviation: 251.8 ms; mean RT during GRAM: 619.4 ms, standard deviation: 224.4 ms; mean EMO–GRAM difference: 95.15 ms, standard deviation: 117.01 ms; Wilcoxon test: $p < 0.0001$). The mean RT during EMO and GRAM was significantly correlated across subjects, indicating that the slow-responders at either task tended to be the same participants ($R^2 = 0.78$). Age was not significantly correlated with RTs during either EMO ($p = 0.1$) or GRAM ($p = 0.5$), whereas education levels were significantly negatively correlated with RTs during EMO ($R^2 = 0.1$, $p = 0.02$) and GRAM ($R^2 = 0.09$, $p = 0.03$), higher level of education being associated with shorter response times.

Regarding the strategies employed by the participants during the EMO task, the most frequently encountered were the analysis of prosodic (92.2%) and lexical (90.2%) emotional cues. Use of any of the three theory of mind strategies, however, was reported by a majority (84.3%) of participants, with a clear dominance of the "social knowledge" item (speaker perspective: 21.6%, addressee perspective: 31.4%, social knowledge: 78.4%).

The average number of recognized sentences after the functional imaging session (out of 12) was higher during EMO than during GRAM (mean number of recognized EMO sentences: 9.02, range: 7–11; mean number of recognized GRAM sentences: 8.4, range: 5–10; mean EMO–GRAM difference: 0.59, standard deviation: 1.43, Wilcoxon test: $p = 0.006$). This indicates that the participants recognized the emotional material of the EMO task more accurately than the neutral material of the GRAM task.

*Imaging data*

*Maps of the event-related response during EMO*

The average time-course maps of the significant BOLD responses to EMO events compared with the baseline (beep detection) are shown in Fig. 1. Hemodynamic activity was first detected at $t = 3$ s post-stimulus, in the Heschl's gyrus and to a lesser extent in the IFG (particularly on the left), precentral gyrus and anterior ventral mPFC. The regions significantly activated at $t = 5$ s further included the superior temporal gyrus, from the temporal pole to the angular gyrus, the medial superior frontal gyrus, from the superior rostral paracingulate area to the supplementary motor area (with a separate patch of ventromedial prefrontal cortex), the calcarine fissures of both hemispheres and the left motor cortex.

Significant activity was also detected at $t = 5$ and 7 s in the posterior cingulate region. In the basal ganglia, activity was seen in the thalamus and *amygdalae* at $t = 5$ s and in the putamen at $t = 7$ s. A late activity (at $t = 11$ s) was observed in the amygdala, right STG, anterior and superior insula. The sensorimotor and premotor regions of the superior parts of the left central, precentral and postcentral sulci and SMA were also active at 11 s, in line with the delayed right-hand motor response relative to the onset of the stimuli.

*EMO minus GRAM comparison*

The EMO minus GRAM contrast (comparison at $t = 7$ s versus $t = 1$ s, summarized in Table 1 and Fig. 2, $p < 0.05$ corrected for multiple comparisons) revealed activations spanning the medial part of the superior frontal gyrus (SFG, Fig. 2), near its medial convexity, and bilaterally. A series of local maxima was also detected along the superior temporal sulcus (STS), from the temporal pole to its posterior angulation, in both hemispheres. These activations covered part of the angular gyrus, with a greater extent in the left hemisphere, and with a stronger more posterior activation peak (24, in the TPJ) which had no counterpart in the right hemisphere. In the frontal lobe, activations were observed in the left and right *pars triangularis* (Fig. 2), along the horizontal and vertical rami, extending inwards up to the circular sulcus of the insula.

In both hemispheres, inferior frontal activations also covered the area at the junction of the precentral sulcus with the inferior frontal sulcus, although local maxima for this region were not present in the left hemisphere. The left and right *amygdalae* (Fig. 2) were more active during EMO than GRAM, at $t = 7$ s compared with $t = 1$ s.

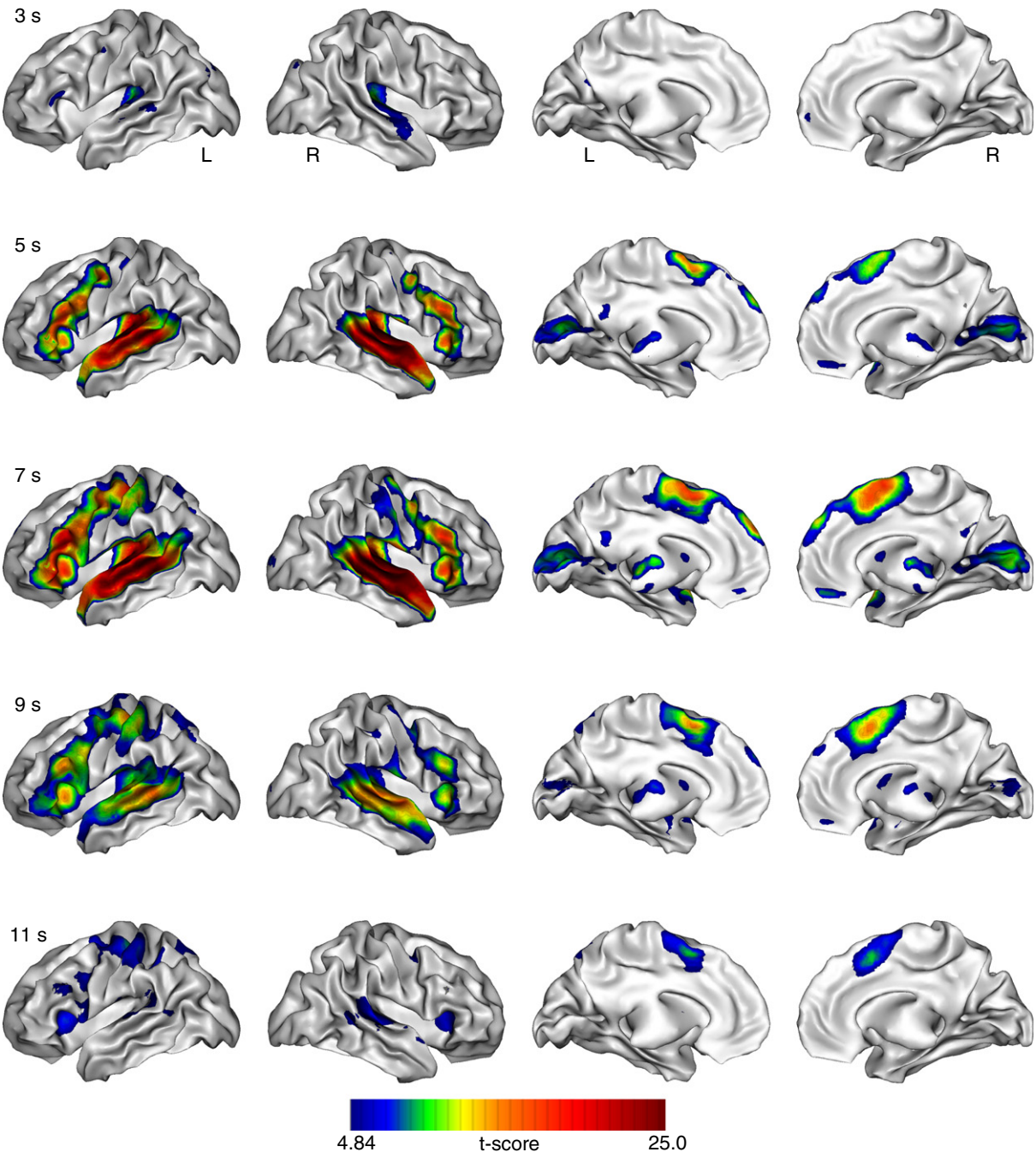


**Fig. 1.** Significant activations at several post-stimulus times during the EMO task, mapped onto the inflated white-matter surface of a "representative brain" in the stereotaxic space ($p<0.05$, FWE). Hotter shades indicate more significant activity (larger t-value) on average at 3, 5, 7, 9 or 11 s after trial onset compared with baseline.

Lastly, in the cortex of the medial-wall, additional foci were detected in the cingulate sulcus, rostral sulcus (region of the gyrus rectus, ventral mPFC), and subparietal sulcus, in the posterior cingulate/inferior precuneus region.

The activation peak in the cingulate sulcus was associated with a lesser deactivation during EMO than during GRAM.

*Network decomposition*

The application of replicator dynamics on the average EMO correlation matrix segmented 8 different networks (Table 1, Fig. 2). This correlation matrix, ordered according to this network decomposition (heatmap), is shown in supplementary Fig. 1. The frequencies (unthresholded membership values) that the individual replicators reached following each iteration of the replicator process are shown in Supplementary Fig. 2.

The 1st iteration selected STS peaks bilaterally, with the exception of those located in the temporal pole and posterior end. Henceforward, this network will be referred to as the *STS* network (in yellow in Fig. 2 and other figures).

The 2nd network consisted of the lateral inferior frontal peaks of both hemispheres. We labeled it as the *Frontal* network (in green).

The 3rd network corresponded to an essentially *Medial* network (in blue): the subparietal sulcus peak (27, pCC) and dorsal medial SFG peaks (mPFC) were associated with a left angular gyrus peak (TPJ peak, 24) and a ventral mPFC peak (30).

The 4th network gathered angular gyrus / STS termination ROIs bilaterally, and was labeled *Angular* network (in cyan).

**Table 1**
Stereotactic coordinates of activation peaks during the EMO task compared with the GRAM task, at t = 7 s versus t = 1 s post-stimulus time. The coordinates are given in the space of the T80 TVS template which is based on the MNI space. Only the local maxima distant by at least 8 mm were kept. The color code matches that of the dots representing the activation peaks in Fig. 2. STS: superior temporal sulcus. STG: superior temporal gyrus. SFG: superior frontal gyrus. The classification of the peaks according to the replicator dynamics at both levels is given.

| Rank | Region | X | Y | Z | t-value | Replicator dynamics | |
|---|---|---|---|---|---|---|---|
| | | | | | | 1st level | 2nd level |
| 9 | Anterior STS R | 56 | −2 | −16 | 9.78 | 1 | |
| 10 | Deep middle STS R | 54 | −16 | −6 | 9.56 | 1 | |
| 12 | Anterior STS L | −52 | 2 | −16 | 9.13 | 1 | |
| 14 | Deep posterior STS R | 50 | −34 | 4 | 8.86 | 1 | 1 |
| 15 | Anterior STS L | −56 | −4 | −12 | 8.80 | 1 | |
| 16 | Deep middle STS L | −54 | −16 | −6 | 8.75 | 1 | |
| 19 | Deep posterior STS L | −52 | −40 | 4 | 7.84 | 1 | |
| 23 | Lat. Posterior STS R | 64 | −34 | 4 | 6.48 | 1 | |
| 2 | Sylv. Hor. ramus L | −48 | 32 | −2 | 12.54 | 2 | |
| 5 | Sylv. Vert. ramus L | −48 | 24 | 6 | 11.29 | 2 | |
| 6 | Sylv. Hor. ramus R | 46 | 28 | −6 | 10.81 | 2 | |
| 17 | Pars triangularis R | 58 | 30 | 14 | 8.16 | 2 | 1 |
| 18 | Insula / Sylv. vert. Ramus R | 36 | 26 | −2 | 8.05 | 2 | |
| 25 | Post. inferior frontal sulcus R | 46 | 20 | 26 | 6.38 | 2 | |
| 26 | Post. inferior frontal sulcus R | 54 | 26 | 24 | 6.05 | 2 | |
| 1 | Med. SFG L | −6 | 56 | 34 | 13.78 | 3 | |
| 3 | Med. SFG R | 6 | 54 | 36 | 11.35 | 3 | |
| 13 | Med. SFG R | 6 | 58 | 24 | 9.08 | 3 | |
| 24 | Angular gyrus L | −42 | −60 | 26 | 6.48 | 3 | 2 |
| 27 | Subparietal sulcus L | −4 | −50 | 28 | 6.05 | 3 | |
| 30 | Superior rostral sulcus L | −2 | 46 | −12 | 5.73 | 3 | |
| 28 | Angular R | 68 | −46 | 16 | 6.00 | 4 | |
| 32 | Angular L | −48 | −56 | 18 | 5.67 | 4 | 1 |
| 33 | Angular L | −56 | −54 | 20 | 5.46 | 4 | |
| 36 | Angular R | 64 | −56 | 18 | 5.13 | 4 | |
| 4 | Polar STS R | 50 | 16 | −24 | 11.29 | 5 | |
| 8 | Polar STS L | −48 | 12 | −28 | 10.43 | 5 | 1 |
| 22 | Temporal Pole R | 50 | 24 | −28 | 6.48 | 5 | |
| 29 | Ant. planum polare R | 46 | 22 | −16 | 5.94 | 5 | |
| 7 | Med. SFG L | −6 | 32 | 54 | 10.65 | 6 | |
| 11 | Med. SFG L | −2 | 42 | 46 | 9.35 | 6 | 2 |
| 21 | Supp Motor Area R | 2 | 14 | 62 | 6.70 | 6 | |
| 20 | Amygdala L | −20 | −8 | −12 | 6.86 | 7 | |
| 31 | Amygdala R | 22 | −6 | −14 | 5.67 | 7 | - |
| 34 | Cingulum L | −8 | 48 | 16 | 5.35 | 8 | |
| 35 | Superior rostral sulcus L | −2 | 38 | −16 | 5.30 | 8 | 2 |
| 37 | Post. STG lateral face, R | 70 | −34 | 14 | 5.08 | - | 1 |

The 5th network was composed of the 4 most anterior temporal activation peaks (labeled *Temporo-polar* network, in orange).

The 6th network was composed of the SMA peak and two anterior peaks of the medial superior frontal gyrus (labeled *SMA* network, purple color).

The 7th network gathered the two *amygdalae* (*amygdalar* network, in magenta).

The 8th and last network was composed of the cingulate gyrus cluster and more posterior ventromedial peak (cingulate network, in red). The remaining right lateral STG peak (number 37, and least significant) was left alone at this last iteration, but was subsequently treated as a 9th network (*STG*, in gray).

At the 2nd level of analysis (i.e. when regrouping the nine networks on the basis of an averaged correlation matrix), the replicator dynamics analysis segmented two network systems.

The 1st iteration selected the Frontal, STS, Temporo-polar and Angular networks and the lone STG peak (37). It was labeled the *Perisylvian* system.

The 2nd iteration gathered the three medial networks (Medial, SMA, Cingulate), and will be referred to as the *Medial* system. The Amygdalar network was singled out during this 2nd iteration.

The ordered between-network correlation matrix is given in Supplementary Fig. 3.

### *Multidimensional scaling*

The configuration of the 37 foci in the functional space defined by the multidimensional scaling analysis can be seen in Fig. 3. The 1st axis separated the medial wall peaks from the rest of the temporal and frontal foci. The first exception to this rule was the left TPJ peak of the Medial network (peak number 24), which was found near the Medial network peaks, in agreement with the replicator dynamics analysis. The second was the SMA peak (21), which was positioned with the Frontal network peaks. An inconsistency with the replicator dynamics thus was apparent at the level of the SMA network, which was a late network (6th).

The 2nd and 3rd axes differentiated the Frontal and STS network peaks, and the *amygdalae* from the other foci. The two amygdalar peaks (7th network) were situated closely in the functional space, but at a far distance from all the other sampled regions. This is consistent with the outcome of both the 1st and 2nd level replicator dynamics analyses.

The 8th cingulate network was not supported by the MDS solution, since a large distance separated the two peaks. Finally, the lateral right STG peak (37), which was not classified by the replicator dynamics analysis, was found near the STS network in the MDS solution.

### *Within-network partial correlations*

Homotopic inter-hemispheric partial correlations were found between peaks 1 and 3 in the medial SFG; 2 and 6 and 5 and 17 in the in the IFG; 4 and 8 in the aTL; 9 and 15, 10 and 16, 14 and 19 in the STS; and 28 and 33 in the posterior STS/angular gyrus (Fig. 4). No heterotopic inter-hemispheric correlation was observed at this threshold.

Within the 3rd network (medial, in blue), partial correlations were observed between the left angular gyrus (24) and the subparietal sulcus (27) foci, indicating a specific association between the two regions. The subparietal sulcus was also related with the right medial SFG (13) and rostral sulcus (30).

More generally, strong partial correlations were observed between closely located ROIs.

### *Average event-related responses*

The average event-related responses (across events and subjects) during the EMO and GRAM tasks for all peaks, sorted according to the 1st and 2nd level networks from the replicator dynamics analysis, are given in Figs. 5, 6 and 7. Overall, the regions composing the Perisylvian 2nd level system were already active during the GRAM task. In these regions, the event-related response peaked earlier during the GRAM task than during the EMO task (see Supplementary Table 1 for statistical analysis). Within the Perisylvian system, the highest level of activity was found in the STS network (yellow in figures), even though the EMO–GRAM difference was often larger in the Frontal network (green in figures). Accordingly, the STS was the most coherent network, and was the first to be extracted by the replicator dynamics. Within each network, large differences in terms of the amplitude of the response could be seen (Fig. 5).

In the Medial 2nd level system (Fig. 6), the EMO event-related BOLD responses were smaller than in the STS or Frontal networks. During the GRAM task, no response, or a slight negative BOLD variation was seen in these regions. Thus, a BOLD response was present only during EMO in these regions (at the sole exception of the SMA). At a descriptive level, this response returned more rapidly to the level observed at t = 1 s (11 s against 15 s in the frontal or temporal network).

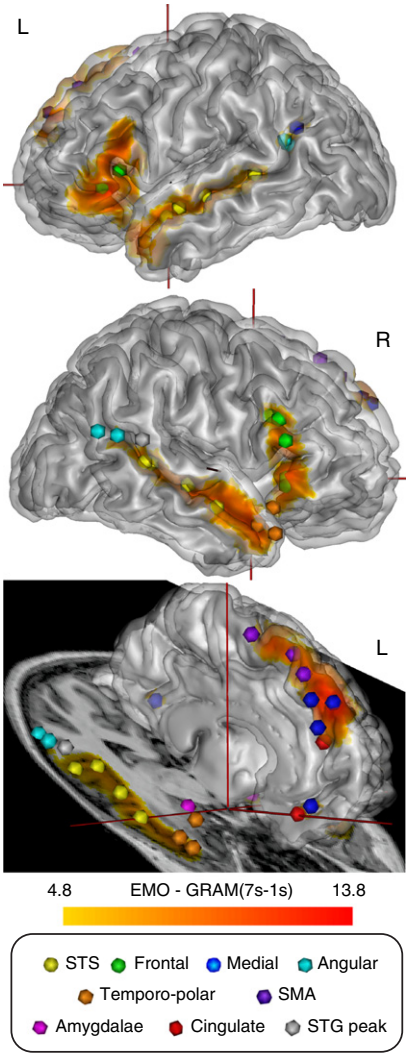


**Fig. 2.** Group activation map of the EMO–GRAM contrast, at t = 7 s versus t = 1 s, overlaid onto a "representative" brain, with a semi-transparent cortical surface and a solid white-matter surface. Hotter shades indicate more significant activity in EMO as compared with GRAM (larger t-values). The colored dots represent the local maxima of the activation map, which defined the regions-of-interest for the functional connectivity analysis. The colors indicate the network-memberships resulting from this analysis, as indicated in the capsule. The oblique plane passes through the superior temporal sulcus and *amygdalae*.

In the *amygdalae*, a consistent response, beginning after 3 s and peaking at 7 s was observed during the EMO task. A response of smaller amplitude was also seen during the GRAM task, between 3 s and 9 s post-stimulus (Fig. 7).

*Summary of the results*

The most salient features of the pattern of functional connectivity between the regions active during emotional sentence classification relative to grammatical judgment are: 1) a separation of the Medial cortical regions from the Perisylvian cortical regions, at the notable exception of a peak in the left TPJ; and 2) a clear segregation between the *amygdalae* and the cortical activation foci (Fig. 3). The bilateral *amygdalae* were grouped together by all connectivity analyses, but were functionally distant from all other regions, and thus formed a third system, in addition to the two other Medial and Perisylvian cortical systems.

At a finer scale, the Perisylvian system was itself composed of several different networks of inferior Frontal, STS, Angular and Temporo-polar regions. The anatomical organization of these four networks was strikingly bilateral. Accordingly, we observed homotopic inter-hemispheric partial correlations in all these networks (Fig. 4). These Perisylvian networks were active to a lesser extent during GRAM, with a shorter time-to-peak during this task compared with EMO.

The Medial system was composed of three networks, with a main Medial network gathering the left TPJ and bilateral dorsal mPFC, along with the ventral mPFC and the pCC. The regions in this network were not active during GRAM. The different functional connectivity analyses we performed did not yield convergent results on the exact composition of the two other networks (SMA and Cingulate).

Finally, the left angular gyrus hosted spatially proximate ROIs belonging to both Perisylvian and Medial systems.

## Discussion

*EMO–GRAM contrast: comparison with the previous study*

The results of our previous study at 1.5 T (Beaucousin et al., 2007) were both replicated and extended. Both reports agree on the relative activations of the dorsal mPFC and pre-SMA, left posterior STS, bilateral inferior frontal gyrus and parts of the right STS during EMO compared with GRAM. In the present study, the STS activations were much more extended, and activations of the *amygdalae*, pCC and ventral mPFC regions were detected during EMO compared with GRAM, whereas they were not in the EMO minus GRAM contrast of the former report. Several methodological improvements were brought in the present study: we used two runs of the actor-voice EMO and GRAM task, an event-related design (compared with a block-design), a sample of 51 participants (compared with 23), and a more recent 3 T scanner (Han and Talavage, 2011; Krasnow et al., 2003; van der Zwaag et al., 2009). We assume that this accounts for the detection of significant activations in the *amygdalae*, pCC and ventral mPFC.

Another difference between these studies study is that, at a behavioral level, the RTs were longer than in the previous study (for EMO: 714 ms against 447 ms in the former study; for GRAM: 619 ms against 385 ms). This is explained by the fact that a longer response delay was allowed in the present slow event-related study (3 s versus 1 s in the former block-design study). The accuracy was also higher in the present study (97.5% at both tasks in the present study, against 84% and 87% respectively for EMO and GRAM with actor voice in the former), suggesting a speed-accuracy trade-off. The same effect of task on RTs, namely, significantly shorter RTs during the GRAM task, was observed in both studies.

*Multiple networks sustain emotional speech comprehension*

We had made the hypothesis that several different networks, with a network gathering the regions frequently co-activated in ToM studies; a network associated with emotional prosody, including the *amygdala* and right STS areas; and a network involving the IFG, which could process emotional words.

The analyses successfully extracted a network of putative ToM areas, validating the 1st hypothesis. In line with this interpretation, in the debriefing interviews, 84% of the participants reported to have used either one of several mind-theoretical strategies, such as self-projections in the place of the sender or receiver and, chiefly, using personal social knowledge. The 2nd hypothesis was also supported by the identification of a bilateral STS network including regions associated with the processing of language and prosody. In contrast to what we had expected, the activity of the *amygdalae* was not associated with that of any other cortical area. The present study thus provided no evidence of a neural interplay between this key emotional region and the right STS regions sensitive to emotional prosody. The 3rd hypothesis found support in the fact that a bilateral

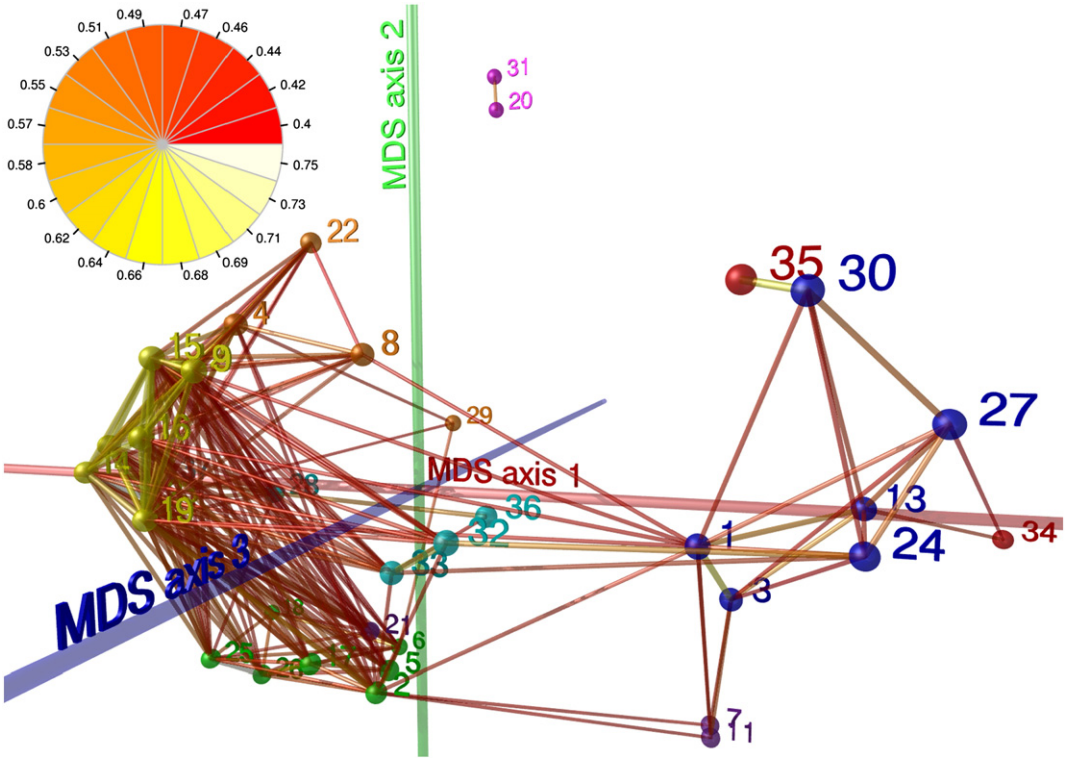


**Fig. 3.** Multidimensional scaling (3 dimensions), applied to the average "functional distance matrix" of the EMO task. The dots of the 37 ROIs are colored according to the classification obtained via the replicator dynamics analysis. The ROIs are numbered according to their rank (Table 1). The rods indicate average correlations above $r = 0.4$ (color coding shown by the pie-chart).

network of inferior Frontal regions was distinguished from both the STS areas and the dmPFC/TPJ network.

This observation of multiple interacting neural systems during emotional speech comprehension is consistent with the outcome of the post-session debriefings. The participants reported using the three strategies: prosodic and lexical cues, as well as social expertise. In addition, several unexpected networks were isolated.

*The Perisylvian system*

This system consisted of four sub-networks, namely the STS, Frontal, Angular and Temporo-polar (Fig. 5). They were activated during the GRAM task versus the baseline, attesting for their involvement in language processing up to the sentence level. The higher activity of this network system during the EMO task may be explained by greater language processing demands compared with the GRAM task. The significantly shorter RTs in the GRAM task indicate that this task was easier to perform than EMO. Indeed, the isolation of the grammatical person of the sentence does not require in-depth semantic analysis. Accordingly, the time-courses of the event-related responses clearly show that the hemodynamic response peaked earlier in the GRAM task in many STS and Frontal network regions (Fig. 5).

These networks were remarkably bilateral. The fact that the left and right homotopes were grouped together systematically is in line with the frequent observation of bilateral coactivations in neuroimaging studies (Toro et al., 2008; Vigneau et al., 2011), and with previous descriptions of stronger homotopic than heterotopic inter-hemispheric correlations during the resting state (Salvador et al., 2005; Stark et al., 2008). The partial correlation analyses further indicate that the inter-hemispheric coherence of these subnetworks is supported in part by specific functional correlations between homotopes. Theoretically, both a predominantly homotopic callosal connectivity and a greater physiological similarity between homologous areas of the two hemispheres should support this homotopic functional connectivity.

Accordingly, no distinction could be made between the left and right STS regions from the functional connectivity analyses. Our previous study using the same tasks with and without affective prosody, however, highlighted a role of the right STS in the treatment of this type of prosody, which is particularly relevant for the EMO task, as indicated by the participants. In this study, the prominent prosody-sensitive right-hemispheric region of the anterior STS identified in the previous study ($x = 56$, $y = -2$, $z = -12$, Beaucousin et al., 2007) was more active during the EMO than during the GRAM task, in line with an effect of emotional prosody. Although the roles of the left and right STS differ, their correlation suggests that these two homotopic regions interact more closely than with other regions during this task, with lateralized but mutually dependent and more synchronous processes. Both ends of the STS have been described as key amodal semantic regions, with a semantic hub in the aTL, a region which is impaired in semantic dementia (Patterson et al., 2007) and a transmodal gateway (Mesulam, 1998) in the posterior STS, corresponding to Wernicke's area. Such amodal regions would link distributed domain-specific representations of a same entity, including verbal labels. Interestingly, these posterior and anterior STS areas were separated from the middle part of the STS by the network decomposition, with bilateral Temporo-polar and Angular networks.

Activations of the aTL, including the level of the STS, have been consistently observed during language comprehension (Mazoyer et al., 1993). The present polar STS (aTL) peaks were found near sentence-specific regions (left: $x = -47$, $y = 17$, $z = -18$; right: $x = 52$, $y = 18$, $z = -20$) involved in the integration of semantic and syntaxic information (Rogalsky and Hickok, 2009). It has been suggested that this region is involved bilaterally in the establishment of "the local coherence of the conceptual units present within the sentence" (Jobard et al., 2007). Ferstl et al. (2008) also evidenced a fairly ubiquitous involvement of the aTL in text comprehension tasks, and suggested that the activity of this region would be associated with the propositionalization of language, the cognitive process by

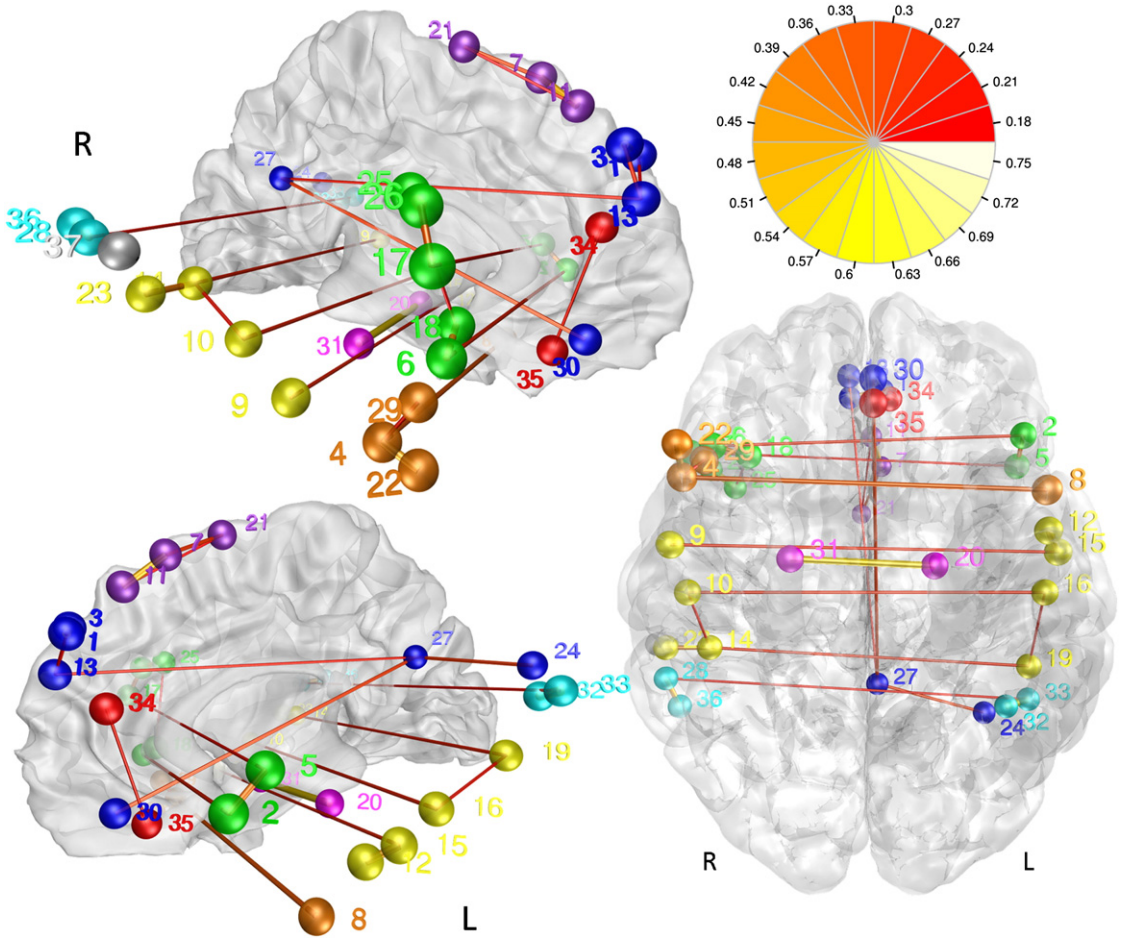


**Fig. 4.** Partial correlations within the 8 networks obtained via the replicator dynamics analysis. The rods indicate partial correlations between two activation foci of a same network exceeding a value of r = 0.175 (color coding shown by the pie-chart in the top right corner). The peaks are numbered according to their rank (Table 1). **Top left**: view from the right. **Bottom left:** view from the left. **Right:** view from below.

which syntactic, prosodic and lexical information are combined in order to extract a semantic or meaning-based representation of the elements of the verbal message (Ferstl, 2007; Ferstl et al., 2008). Furthermore, in that study, the replicator dynamics analysis of the meta-analytical co-activation data on text compared with rest or perceptual baselines isolated the bilateral anterior STS as a part of the dominant network, what suggested a close relation between the left and right aTLs. This is supported by the observed partial correlation between the two aTL homotopes (8 and 4).

The posterior end of the STS (Angular network) is involved in semantic integration at the sentence level and beyond the sentence boundary: as a matter of fact, this network corresponds to the STSp cluster ($x = -50$, $y = -54$, $z = 22$) of a previous meta-analysis of contrasts based on single sentences or groups of sentences (Vigneau et al., 2006). The posterior STS region shows a greater activation during narrative processing compared with sentence processing (Jobard et al., 2007; Xu et al., 2005). The posterior STS region is also active during the production of overt signed or spoken narratives (Braun et al., 2001).

Despite all this evidence in favor of the involvement of these regions in sentence and text processing, the finding that the Temporo-polar and Angular networks were closer to the Perisylvian system than to the Medial system is interesting. The Temporo-polar ROIs are located closely to person-selective regions (left: $x = -47$, $y = 17$, $z = -22$; right: $x = 47$, $y = 17$, $z = -26$ mm) of the lateral anterior temporal lobe (Simmons et al., 2010). At rest, these are included in a wider social cognition network, which is organized around pCC, anterior and posterior temporal and mPFC regions (Gallagher and Frith, 2003; Mar, 2011; Saxe and Kanwisher, 2003). One could have expected both the aTL and all the ROIs of the angular gyrus/posterior STS (peaks 24, 28, 32, 33, 36) to associate with the medial regions in the functional connectivity analyses. Yet, the Temporo-polar network was more correlated with the other Perisylvian networks than with the Medial networks. Different aTL regions may have been active in the two studies, although the stereotaxic coordinates were relatively close. Another possibility is that a single aTL region would exhibit changes in terms of functional connectivity depending on the cognitive context. At the level of the inferior angular gyrus, the network-decomposition suggests the coexistence of two neighboring, but distinct regions, one interacting with the perisylvian language regions (Angular network, peaks 28, 32, 33, 36) and the other interacting with medial brain regions (TPJ, peak 24). This novel finding is consistent with the fact that mPFC and pCC regions accompanied the bilateral TPJ in a comparison of narrative and sentence processing (Xu et al., 2005), and with the fact that the STSp cluster in Vigneau et al. (2006) also included peaks from ToM studies.

### *The Medial system*

Among the 37 active ROIs, the regions classically involved in ToM (TPJ, pCC, mPFC) were aggregated together robustly in the 1st level Medial network, at the notable exception of the aTL regions, as discussed earlier. Other ventral and more posterior dorsal medial regions joined this network at the 2nd level analysis. This medial system showed greater activity only when active inferences on the emotional sentences had to be made, relative to baseline, and not when

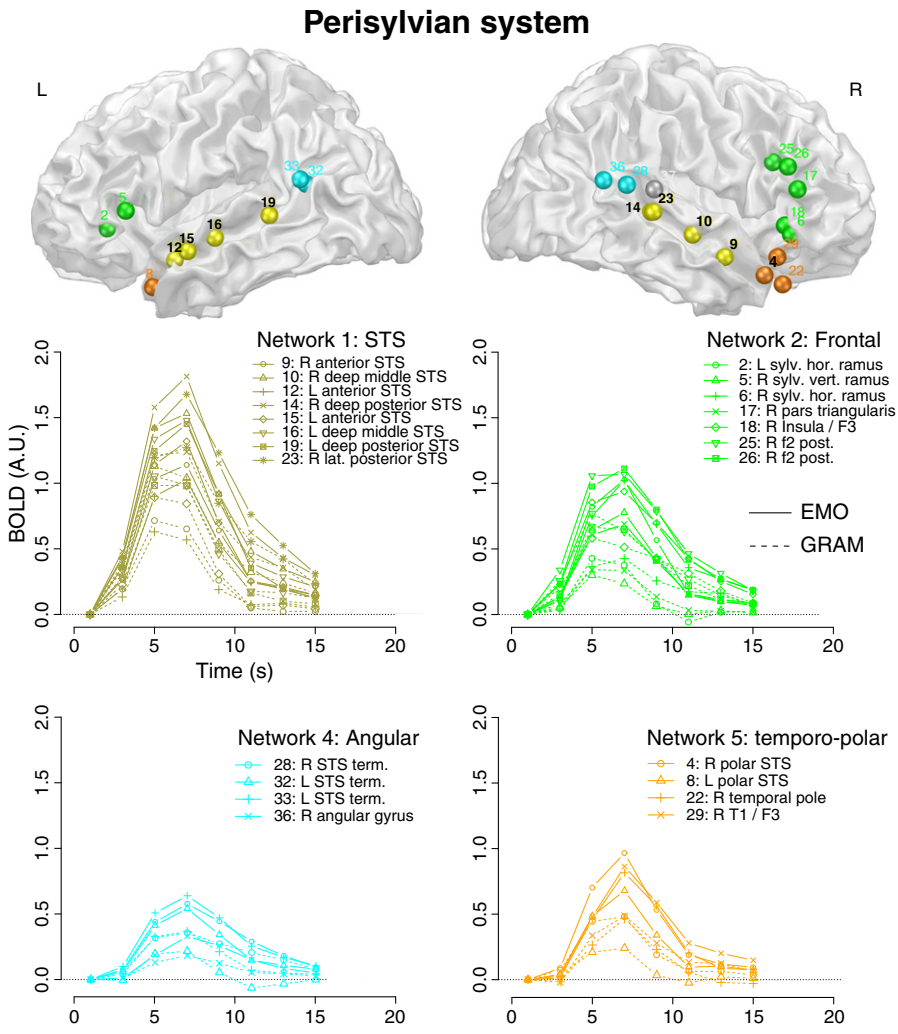


**Fig. 5.** Average event-related BOLD responses for the Perisylvian system (1 to 15 s post-stimulus) during the EMO (solid lines) and GRAM (dashed lines) tasks, for each ROI, grouped according to the results of the replicator dynamical analysis. T = 1 s was used as a baseline.

the 3-choice classification was performed according to the grammatical person of the sentences (Fig. 6). The following discussion will focus on the 1st level Medial network.

Why was this network recruited during the EMO task? We can rule out an effect of prosody: in the previous study (Beaucousin et al., 2007), the absence of emotional prosody did not influence the level of activity of the dorsal mPFC or left TPJ during the EMO task.

Our hypothesis is that this network, and more particularly the mPFC, would be involved in the drawing of emotional inferences on the basis of the sentence meanings. The participants who reported using their social knowledge often reported to have relied on their knowledge of how people generally feel when facing the situations depicted by the sentences, notably for sentences involving commuters being caught in a train strike, a couple breaking-up their romantic relationship, or a person going on a holiday. This corresponds to the kind of inferences made during text comprehension, or during a ToM reasoning. The dorsal mPFC could drive non-automatic inference processes during text comprehension (Ferstl, 2007; Ferstl and von Cramon, 2001, 2002). It is strongly associated with ToM tasks as well (Ferstl and von Cramon, 2002; Frith and Frith, 2006). During a complex strategic reasoning task, the dorsal mPFC is also more active when one's answer is based on a more elaborate prediction of the other people's answers (Coricelli and Nagel, 2009).

The pCC is frequently co-activated with the dorsal mPFC during ToM (Mar, 2011) or text comprehension (Ferstl et al., 2008). This was the case in a PET study in which the brain activations during the processing of stories whose comprehension required prior knowledge in the form of a relevant picture were compared with the activations during usual stories. The results showed that the pCC was associated with the linkage of the non-verbal prior knowledge given to the participants with the incoming verbal information, while building and updating a situation model (Maguire et al., 1999).

In this same PET study, a region of the ventromedial orbitofrontal cortex was correlated with the degree of comprehension, but not with the presence or absence of prior knowledge (Maguire et al., 1999). This region corresponds with the ventral mPFC region of the present study (peaks 30 and 35). Several studies show that this region responds when judgments or inferences are made on social and/or emotional material, as is the case during the EMO task. This region is activated when participants passively view people being hurt intentionally versus accidentally (Decety et al., 2012). It is also recruited, among other nodes of the present Medial network, when processing moral versus non-moral facts about others (Young and Saxe, 2009). Besides, in an experiment involving hot reasoning on syllogisms with emotionally salient words, it was only during the hot condition that the ventromedial prefrontal cortex was more active during reasoning than baseline (Goel and Dolan, 2003).

The Medial-network left TPJ peak coordinates correspond with the centre-of-mass of the semantic AG cluster ($x = -45$, $y = -68$, $z = 26$, medial to the sentence-level STSp cluster in the meta-analysis by

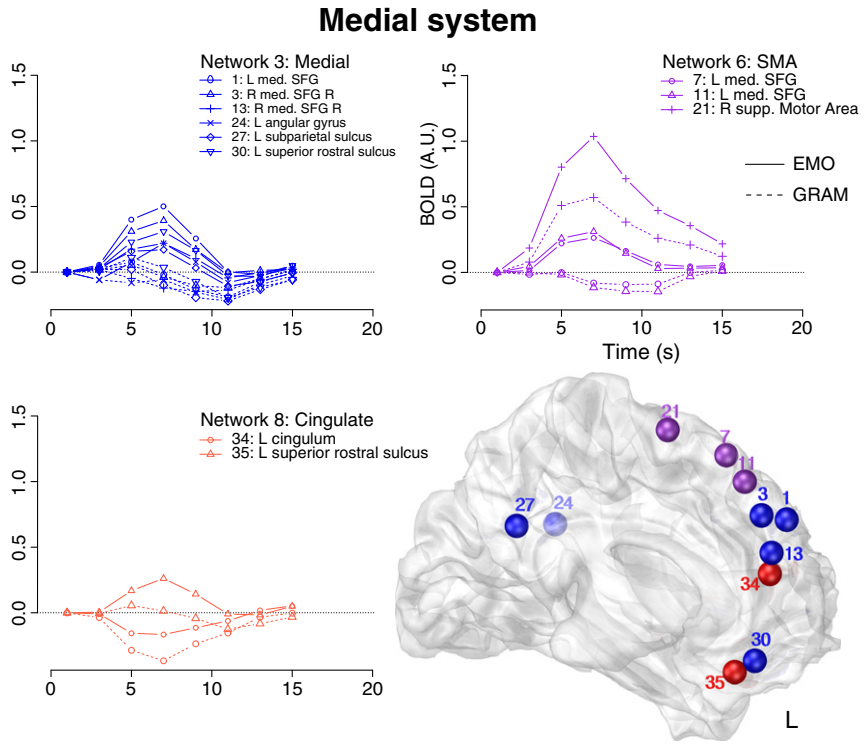


**Fig. 6.** Average event-related BOLD responses for the Medial system (1 to 15 s post-stimulus) during the EMO (solid lines) and GRAM (dashed lines) tasks, for each ROI, grouped according to the results of the replicator dynamical analysis. T = 1 s was used as a baseline.

Vigneau et al., 2006), but this ROI was not classified with the perisylvian regions by the present functional connectivity analyses. Accordingly, the impact of a lesion of this region is not limited to the verbal domain: lesions of the posterior STG/angular gyrus (Wernicke's area) impair the processing of both speech and meaningful non-verbal auditory information, such as environmental sounds (Saygin et al., 2003). The left TPJ has also been linked with the ability to build complex representations of the beliefs of others, using both functional imaging and the lesion approach (Samson et al., 2004; Saxe and Kanwisher, 2003), and with both verbal and non-verbal tasks (Castelli et al., 2000; Samson et al., 2004). It was shown recently that the semantic system and the default-mode network, which is also organized around mPFC, precuneus and inferior parietal lobe nodes and may support planning or problem solving during rest, overlap in the angular gyrus (Binder et al., 1999; Seghier et al., 2010).

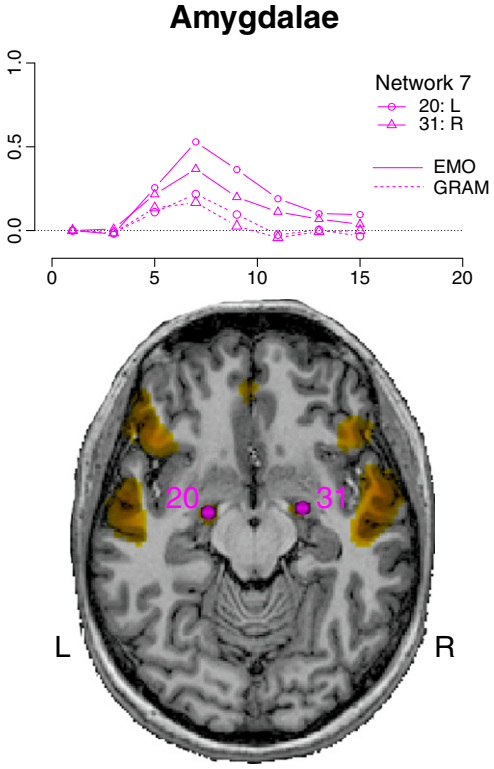


**Fig. 7.** Average event-related BOLD responses for the Medial system (1 to 15 s post-stimulus) during the EMO (solid lines) and GRAM (dashed lines) tasks, for each ROI, grouped according to the results of the replicator dynamical analysis. T = 1 s was used as a baseline.

*The amygdalae*

In the previous study, the activity of the *amygdala* was significantly influenced by the presence of emotional prosody, but not by the task: the activity of the *amygdala* was greater in the EMO task with the actor voice, compared with the same task with the software voice, but was similar in the EMO and GRAM tasks with the actor voice. By contrast, the present study reveals a task effect (EMO higher than GRAM with the actor voice) on the activity of the *amygdalae*. This is, however, consistent with previous observations of amygdalar activity during the processing of narratives (Xu et al., 2005), and of a sensitivity of the *amygdalae* to prosody (Wiethoff et al., 2009). Indeed, the EMO stimuli contained emotional prosody against neutral prosody during GRAM. This differential involvement of the amygdala in the EMO and GRAM tasks with actor voices may simply have been undetected in the previous study. A small amygdalar response, however, was already present in the GRAM task (Fig. 7), what suggests that a same process active during GRAM might have been enhanced during EMO. For instance, participants may monitor their performance during the tasks, what could trigger emotional reactions. A potential behavioral correlate of the amygdalar activation lies in the fact that, at the post-session interview, the EMO task stimuli were recalled with a better accuracy than the GRAM stimuli. The amygdala

has been shown to be involved in such enhancement of memory processes by emotions (Cahill et al., 1996).

In terms of resting-state functional connectivity, the *amygdala* region is known to be correlated positively with the ventromedial prefrontal cortex, the lateral temporal lobe and the inferior frontal gyrus, but negatively with the dorsomedial prefrontal cortex (Roy et al., 2009). In the present analyses, however, the *amygdalae* were not associated with any of these regions, at both iterations of the replicator dynamics algorithm. If one assumes that functional correlations reflect the degree of interactions between regions, the limited correlation with the other brain regions during the EMO task suggests that the *amygdalae* were integrated poorly in the network of areas more active during the EMO task, and that their actual role in prosodic emotion recognition might be marginal.

## Conclusion

Functional connectivity analyses isolated different cortical network systems, which were functionally integrated during the emotional sentence classification task. Two of these systems also observed during the processing of single sentences (Perisylvian and Medial), might be understood in the light of a model of the neural correlates of discourse processing previously suggested by Xu et al. (2005). Based on their observations, these authors proposed that discourse comprehension and production relied on the integration of a ventro-lateral system and a dorso-medial system, providing respectively the "tools" for language and the "motivation" for their use. Within this frame, the activity of the Medial system observed here did not increase when the task was focused on the grammatical aspects of the sentences, but did change when the focus was shifted onto their emotional content as inferences became useful to perform.

By contrast, the Perisylvian system was already active during the GRAM task, and its activity increased during the EMO task. This may be because a deeper semantic analysis of the sentences is required by the inference-making process driven by the Medial network, in response to task demands. The bilateral Perisylvian system would be involved mostly in paralinguistic (prosody and context) and linguistic (phonological, lexical and syntaxic levels) analyses, and combine them in order to extract units of meaning beyond the word level (i.e. propositions). Conversely, the Medial system would rather operate on such conceptual knowledge produced by the Perisylvian language system and other emotion or action recognition systems, in order to complement the conceptual representations of the stimuli. In the context of this study, such a representation would include the emotional state of the speaker, based on the recalled emotional correlates of the situations depicted by the sentences, and influenced by the tone of the speaker's voice and isolated lexical cues.

Finally, to the extent that active areas belonging to each system were found in close vicinity within the AG (TPJ) during emotional speech classification, our results suggest that this amodal semantic region might constitute an interface between these two Perisylvian and Medial, linguistic and conceptual, systems.

## Acknowledgments

The authors thank Virginie Beaucousin, Marie-Renée Turbelin, Guy Perchey, Emmanuel Mellet and Elise Leroux for data collection and Mickaël Naveau, Marc Joliot and Nicolas Delcroix for their help with data analysis, as well as Gaël Jobard for helpful discussions on the manuscript. This research was funded by the Agence Nationale de la Recherche (ANR-05-NEURO-34-01).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.neuroimage.2012.03.073.

## References

Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, B., Tzourio-Mazoyer, N., 2007. FMRI study of emotional speech comprehension. Cereb. Cortex 17, 339–352.
Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Rao, S.M., Cox, R.W., 1999. Conceptual processing during the conscious resting state. A functional MRI study. J. Cogn. Neurosci. 11, 80–95.
Braun, A.R., Guillemin, A., Hosey, L., Varga, M., 2001. The neural organization of discourse: An H215O-PET study of narrative production in English and American sign language. Brain 124, 2028–2044.
Cahill, L., Haier, R.J., Fallon, J., Alkire, M.T., Tang, C., Keator, D., Wu, J., McGaugh, J.L., 1996. Amygdala activity at encoding correlated with long-term, free recall of emotional information. Proc. Natl. Acad. Sci. 93, 8016–8021.
Castelli, F., Happé, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12, 314–325.
Coricelli, G., Nagel, R., 2009. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proc. Natl. Acad. Sci. 106, 9163–9168.
Decety, J., Michalska, K.J., Kinzler, K.D., 2012. The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. Cereb. Cortex 22 (1), 209–220.
Desai, R.H., Binder, J.R., Conant, L.L., Seidenberg, M.S., 2010. Activation of sensory-motor areas in sentence comprehension. Cereb. Cortex 20, 468–478.
Ferstl, E.C., 2007. The functional neuroanatomy of text comprehension: what's the story so far? In: Schmalhofer, F., Perfetti, C.A. (Eds.), Higher level language processes in the brain: inference and comprehension processes. Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 53–102.
Ferstl, E.C., von Cramon, D.Y., 2001. The role of coherence and cohesion in text comprehension: an event-related fMRI study. Cogn. Brain Res. 11, 325–340.
Ferstl, E.C., von Cramon, D.Y., 2002. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? Neuroimage 17, 1599–1612.
Ferstl, E.C., Neumann, J., Bogler, C., von Cramon, D.Y., 2008. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. Hum. Brain Mapp. 29, 581–593.
Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S.J., Frith, C.D., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57, 109–128.
Friston, K.J., Frith, C.D., Fletcher, P., Liddle, P.F., Frackowiak, R.S., 1996. Functional topography: multidimensional scaling and functional connectivity in the brain. Cereb. Cortex 6, 156–164.
Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. Neuron 50, 531–534.
Gallagher, H.L., Frith, C.D., 2003. Functional imaging of "theory of mind". Trends Cogn. Sci. 7, 77–83.
Gallagher, H.L., Happé, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D., 2000. Reading the mind in cartoons and stories: an fMRI study of "theory of mind" in verbal and nonverbal tasks. Neuropsychologia 38, 11–21.
Goel, V., Dolan, R.J., 2003. Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. Neuroimage 20, 2314–2321.
Han, K., Talavage, T.M., 2011. Effects of combining field strengths on auditory functional MRI group analysis: 1.5 T and 3 T. J. Magn. Reson. Imaging 34, 1480–1488.
Jobard, G., Vigneau, M., Mazoyer, B., Tzourio-Mazoyer, N., 2007. Impact of modality and linguistic complexity during reading and listening tasks. Neuroimage 34, 784–800.
Kintsch, W., 1988. The role of knowledge in discourse comprehension: a construction-integration model. Psychol. Rev. 95, 163–182.
Krasnow, B., Tamm, L., Greicius, M., Yang, T., Glover, G., Reiss, A., Menon, V., 2003. Comparison of fMRI activation at 3 and 1.5 T during perceptual, cognitive, and affective processing. Neuroimage 18, 813–826.
Lohmann, G., Bohn, S., 2002. Using replicator dynamics for analyzing fMRI data of the human brain. IEEE Trans. Med. Imaging 21, 485–492.
MacSweeney, M., Capek, C.M., Campbell, R., Woll, B., 2008. The signing brain: the neurobiology of sign language. Trends Cogn. Sci. 12, 432–440.
Maguire, E.A., Frith, C.D., Morris, R.G.M., 1999. The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. Brain 122, 1839–1850.
Mangin, J.-F., Riviere, D., Cachia, A., Duchesnay, E., Cointepas, Y., Papadopoulos-Orfanos, D., Scifo, P., Ochiai, T., Brunelle, F., Regis, J., 2004. A framework to study the cortical folding patterns. Neuroimage 23, S129–S138.
Mar, R.A., 2011. The neural bases of social cognition and story comprehension. Annu. Rev. Psychol. 62, 103–134.
Martin, A., Chao, L.L., 2001. Semantic memory and the brain: structure and processes. Curr. Opin. Neurobiol. 11, 194–201.
Mazoyer, B.M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., Mehler, J., 1993. The cortical representation of speech. J. Cogn. Neurosci. 5, 467–479.
Mesulam, M.M., 1998. From sensation to cognition. Brain 121, 1013–1052.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. 8, 976–987.
Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. J. Neurosci. 30, 10127–10134.
Rogalsky, C., Hickok, G., 2009. Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. Cereb. Cortex 19, 786–796.
Roy, A.K., Shehzad, Z., Margulies, D.S., Kelly, A.M.C., Uddin, L.Q., Gotimer, K., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2009. Functional connectivity of the human amygdala using resting state fMRI. Neuroimage 45, 614–626.

Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., Bullmore, E., 2005. Neurophysiological architecture of functional magnetic resonance images of human brain. Cereb. Cortex 15, 1332–1342.
Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporoparietal junction is necessary for representing someone else's belief. Nat. Neurosci. 7, 499–500.
Saxe, R., 2006. Uniquely human social cognition. Curr. Opin. Neurobiol. 16, 235–239.
Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". Neuroimage 19, 1835–1842.
Saygin, A.P., Dick, F., Wilson, S.M., Dronkers, N.F., Bates, E., 2003. Neural resources for processing language and environmental sounds: evidence from aphasia. Brain 126, 928–945.
Seghier, M.L., Fagan, E., Price, C.J., 2010. Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. J. Neurosci. 30, 16809–16817.
Simmons, W.K., Reddish, M., Bellgowan, P.S.F., Martin, A., 2010. The selectivity and functional connectivity of the anterior temporal lobes. Cereb. Cortex 20, 813–825.
Stark, D.E., Margulies, D.S., Shehzad, Z.E., Reiss, P., Kelly, A.M.C., Uddin, L.Q., Gee, D.G., Roy, A.K., Banich, M.T., Castellanos, F.X., Milham, M.P., 2008. Regional variation in interhemispheric coordination of intrinsic hemodynamic fluctuations. J. Neurosci. 28, 13754–13764.
Toro, R., Fox, P.T., Paus, T., 2008. Functional coactivation map of the human brain. Cereb. Cortex 18, 2553–2559.
van der Zwaag, W., Francis, S., Head, K., Peters, A., Gowland, P., Morris, P., Bowtell, R., 2009. fMRI at 1.5, 3 and 7 T: Characterising BOLD signal changes. Neuroimage 47, 1425–1434.
Vigneau, M., Beaucousin, V., Hervé, P.Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. Neuroimage 30, 1414–1432.
Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2011. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing?: insights from a meta-analysis. Neuroimage 54, 577–593.
Wiethoff, S., Wildgruber, D., Grodd, W., Ethofer, T., 2009. Response and habituation of the amygdala during processing of emotional prosody. Neuroreport 20, 1356–1360.
Xu, J., Kemeny, S., Park, G., Frattali, C., Braun, A., 2005. Language in context: emergent features of word, sentence, and narrative comprehension. Neuroimage 25, 1002–1015.
Young, L., Saxe, R., 2009. An fMRI investigation of spontaneous mental state inference for moral judgment. J. Cogn. Neurosci. 21, 1396–1405.
Zwaan, R.A., 1999. Situation models: the mental leap into imagined worlds. Curr. Dir. Psychol. Sci. 8, 15–18.