

FMRI Study of Emotional Speech Comprehension

Virginie Beaucousin¹, Anne Lacheret², Marie-Renée Turbelin³, Michel Morel², Bernard Mazoyer^{1,3} and Nathalie Tzourio-Mazoyer¹

¹Groupe d'Imagerie Neurofonctionnelle, UMR6194, Centre National de la Recherche Scientifique/CEA/Universités Caen et Paris 5, France, ²Centre de Recherches Inter-langues sur la Signification en Contexte, FRE 2805, Centre National de la Recherche Scientifique/Université Caen and ³IRM CHU Caen, Institut Universitaire de France

Little is known about the neural correlates of affective prosody in the context of affective semantic discourse. We used functional magnetic resonance imaging to investigate this issue while subjects performed 1) affective classification of sentences having an affective semantic content and 2) grammatical classification of sentences with neutral semantic content. Sentences of each type were produced half by actors and half by a text-to-speech software lacking affective prosody. Compared with neutral sentences processing, sentences with affective semantic content—with or without affective prosody—led to an increase in activation of a left inferior frontal area involved in the retrieval of semantic knowledge. In addition, the posterior part of the left superior temporal sulcus (STS) together with the medial prefrontal cortex were recruited, although not activated by neutral sentences classification. Interestingly, these areas have been described as implicated during self-reflection or other's mental state inference that possibly occurred during the affective classification task. When affective prosody was present, additional rightward activations of the human-selective voice area and the posterior part of STS were observed, corresponding to the processing of speaker's voice emotional content. Accurate affective communication, central to social interactions, requires the cooperation of semantics, affective prosody, and mind-reading neural networks.

Keywords: emotion, fMRI, language, prosody, theory of mind

Introduction

Emotional verbal communication is a fundamental element of human relationships in which comprehension emerges from the processing of both linguistic and pragmatic information. Whereas linguistic information includes the integration of the meaning of words (semantics) and sentences (syntax), pragmatic information concerns the processing of gestures, facial expressions, and emotional prosody that accompanies the oral expression of language. How do semantic and emotional prosodic contents interact in the brain to complete an accurate comprehension of emotional discourse?

The neural basis of neutral speech understanding is well documented, and numerous reports have led to a quite clear definition of the left hemispheric frontal and temporal areas involved in semantic and syntactic processing (for a review, see Vigneau and others 2006). Regarding the question of the neural basis of emotional speech processing, studies are scarce. Only one study has investigated the semantic integration of emotional discourse at the word level (Beauregard and others 1997), whereas most reports have focused on the neural implementation of emotion conveyed by prosody. A survey of these studies shows that the comprehension of emotional prosody recruits both the right inferior frontal and temporal areas (Mitchell and

others 2003) together with homologous leftward regions that are known to process the linguistic aspects of language (Wildgruber and others 2002, 2004, 2005; Kotz and others 2003; Grandjean and others 2005). Such evidence suggests the involvement of syntactico-semantic areas during emotional prosodic processing and questions the specificity of right temporal areas for emotional prosodic processing, a conclusion based on the observations of aprosodic patients (Ross 1981).

However, before a definite conclusion can be drawn, the possible impact of the paradigms used in the functional imaging studies on the results obtained must be considered. In these reports, subjects were generally presented with auditory stimuli having a semantic content inconsistent with emotional prosody (e.g., unintelligible stimuli [Kotz and others 2003], sentences constructed with pseudowords [Grandjean and others 2005], or sentences with neutral content [Wildgruber and others 2002, 2004, 2005] spoken with emotional prosody). Such a paradigm, designed to remove semantic processing from the cognitive task, could have paradoxically led to an increase in the semantic demand. This proposed effect is supported by the observation that the more the speech is unintelligible, the more the activity in the left perisylvian semantic areas increases (Meyer and others 2002; Kotz and others 2003).

The effect could also attest the existence of a close cooperation between semantic and prosodic systems necessary to perform accurate speech comprehension. As a matter of fact, strong interactions do exist between prosodic and semantic systems during emotional speech comprehension: a drastic reduction in performance occurs in normal volunteers during affective categorization of sentences lacking syllabification but including affective prosody (Lakshminarayanan and others 2003), whereas aprosodic patients dramatically improve their scores during affective speech comprehension to 70% correct answers (CA) when affective sentences include congruent semantic content (Bowers and others 1987).

To achieve the goal of disentangling linguistic and prosodic neural components and to further investigate their relative involvement during affective speech comprehension, we elaborated a functional magnetic resonance imaging (fMRI) paradigm. To uncover the areas processing affective prosody, we compared the affective classification of sentences spoken by actors with the affective classification of sentences produced by Kali, a text-to-speech software that generates sentences from natural spoken syllables, which includes grammatical prosody but is devoid of affective prosody (Morel and Lacheret-Dujour 2001). Sentences produced by actors and Kali were equivalent in terms of affective syntactico-semantic content. In order to explore the neural correlates of affective sentence comprehension, we compared the regions involved in affective classification

of sentences having an affective (i.e., emotional and attitudinal) semantic content with the areas obtained during the grammatical classification of sentences having neutral semantic content. This comparison was operated by means of a conjunction on sentences enounced by actors and Kali to provide the neural network of affective semantic content comprehension independently of the presence of affective prosody.

Materials and Methods

Elaboration of the Stimuli

Sentence Construction

The initial corpus was composed of 120 sentences with emotional semantic content (anger, sadness, or happiness), 120 sentences with attitudinal semantic content (obviousness, doubt, or irony), and 80 neutral sentences (see Appendix A, for examples of sentences). The construction of the sentences was identical for all types of sentences: length, word frequency, and imageability were matched across conditions. The words that composed the sentences consisted of 2 or 3 syllables and were frequent (on a text sample of 100 million words, the words selected appeared at least 2000 times) and highly imageable (scored 5 out of 6) as assessed in the Brulex database (Content and others 1990). Although most of the sentences had a simple syntactic structure, including subject, verb, and complement, 107 sentences have a more complex structure with an additional complement. These more complex sentences were equally represented in each category, corresponding to 5% of the total number of sentences. The length of the sentences was equivalent across conditions (mean number of words per sentence including functional words: 9.2 ± 2 for neutral, 10 ± 2 for happiness, 10.4 ± 2 for sadness, 11.4 ± 3 for anger, 9.1 ± 2 for doubt, 10.2 ± 2 for obviousness, 10.7 ± 2 for irony).

Sentence Stimuli Production

All sentences were recorded twice. They were either enounced by actors with appropriate grammatical and affective prosody or produced by a text-to-speech synthesis that constructs sentences from naturally spoken syllables, that includes grammatical prosody but lacks affective prosody (Kali) (Morel and Lacheret-Dujour 2001). In order to avoid the possible confound of speaker's gender (Wildgruber and others 2002), half of the sentences produced by Kali were pronounced with a female voice whereas the remaining half were produced with a male voice. Similarly, half of the sentences were enounced by an actor and the other half by an actress.

Sentence Selection

A pilot experiment was conducted with 16 subjects (9 men) to select the affective and neutral sentences to be used in the functional study. The criteria were that 1) affective (emotional and attitudinal) sentences enounced by actors had to be accurately classified by all the subjects and 2) affective and neutral sentences produced by Kali had to be understood by all the subjects. Starting with 320 sentences, 180 affective sentences were chosen, 30 per category, including 15 produced by Kali and 15 enounced by actors. In addition, a set of 90 neutral sentences was selected.

Subjects

Twenty-three young healthy volunteers participated in the fMRI study, 11 men and 12 women (23.3 ± 3 years). All were right handed (Edinburgh score = 88.7 ± 13 [Oldfield 1971]), university students (4 years at university ± 2 years) reported French as their mother tongue and were selected as having a typical leftward hemispheric asymmetry on functional images. They had no auditory deficit, and their T_1 -weighted magnetic resonance images were free from abnormalities. All gave informed written consent to the study, which was accepted by our local ethical committee (Comité Consultatif de Protection des Personnes pour la Recherche Biomédicale de Basse-Normandie no. 99/36).

Procedure

Prior to the experiment, subjects were given instructions and training. During the fMRI acquisitions, the subjects were submitted to 8 runs: 4 runs of affective classification and 4 runs of grammatical classification. During the affective classification task, the subjects had to classify the affective sentences that they heard into 1 of the 3 categories. The emotional sentences had to be classified as happy, angry, or sad, whereas attitudinal sentences had to be classified as expressing doubt, irony, or obviousness. Both affective classifications (emotional and attitudinal) were performed in separate runs twice: once with sentences enounced by the actors, that is, with affective prosody (2 runs AffAct) and once with different sentences produced by Kali, that is, devoid of affective prosody (2 runs AffKali).

During the grammatical classification, the subjects had to classify the subject of the sentences according to its form: first, second, or third person. This task was performed on neutral sentences in 2 different runs: in one run the sentences were enounced by actors (GrAct) and in the other they were produced by Kali (GrKali). Both runs were replicated with different sentences to further allow independent statistical comparison with the 4 runs of affective classification.

The subjects were given first the 4 runs containing sentences produced by Kali (2 AffKali and 2 GrKali) to avoid an influence of the affective prosody carried by the sentences enounced by actors. Within each session of 4 runs, pseudorandom presentation was used to avoid a confounding effect of order.

Outside the scanner, subjects answered a postsession questionnaire to determine the strategy they used to perform the affective classification. They were asked whether they used semantic (meaning of the words), syntactic (structure of the sentences), or prosodic cues (intonation) and whether they had rehearsed the sentences produced by Kali or enounced by actors.

Experimental Design and Apparatus

A block design was constructed for the present paradigm. Each of the 8 runs lasted 6 min 36 s and began with 60 s of a control task consisting in the detection of 9 beeps, presented through earphones at random interstimuli intervals. Subjects had to press alternatively 1 of the 3 buttons on a pad. This block was followed by 5 blocks of classification task lasting 34 s alternating with 5 blocks of control task that lasted 32 s. During each block of classification task, subjects listened to 9 sentences per block, each lasting about 3 s (Fig. 1). Answers and response time (RT), limited to 1 s, were collected using a pad with 3 buttons corresponding to the 3 choices proposed to the subjects in each classification task. The answers were assigned to the key following the alphabetic order of the categories to give the subject a mnemonic mean (e.g., in the emotional classification, anger was assigned to key 1, happiness to key 2, and sadness to key 3). Presentation of stimuli and recording of the responses were achieved through a computer equipped with the software SuperLab™ Pro version 2.0 (Cedrus, <http://www.superlab.com/papers>). Because of technical reasons, responses could not be acquired for 3 male participants.

Analysis of Behavioral Data

The percentage of CA and RTs for CA (RT, ms) were recorded during the fMRI acquisition in each classification task. Eight men out of 11 and 12 women were included in this analysis (scores for 3 men were missing due to technical problems). Kolmogorov and Smirnov test was performed to assess whether the variables' distribution differed from normality, which was not the case (CA for neutral sentences $\chi^2 = 3.8$, $P = 0.3$, for affective sentences $\chi^2 = 1.7$, $P = 0.8$; RT for neutral sentences $\chi^2 = 0.7$, $P > 0.99$, for affective sentences $\chi^2 = 0.7$, $P > 0.99$). An analysis of variance (ANOVA) with repeated measures was thus conducted with 2 factors: Task (affective vs. grammatical classification) and Voice (actors vs. Kali). Post hoc comparisons were performed using paired *t*-tests with a Bonferroni correction.

Analysis of Images

Acquisition of Images

Magnetic resonance imaging (MRI) acquisitions were conducted on a GE Signa 1.5-T Horizon Echospeed scanner (General Electric, BUC,

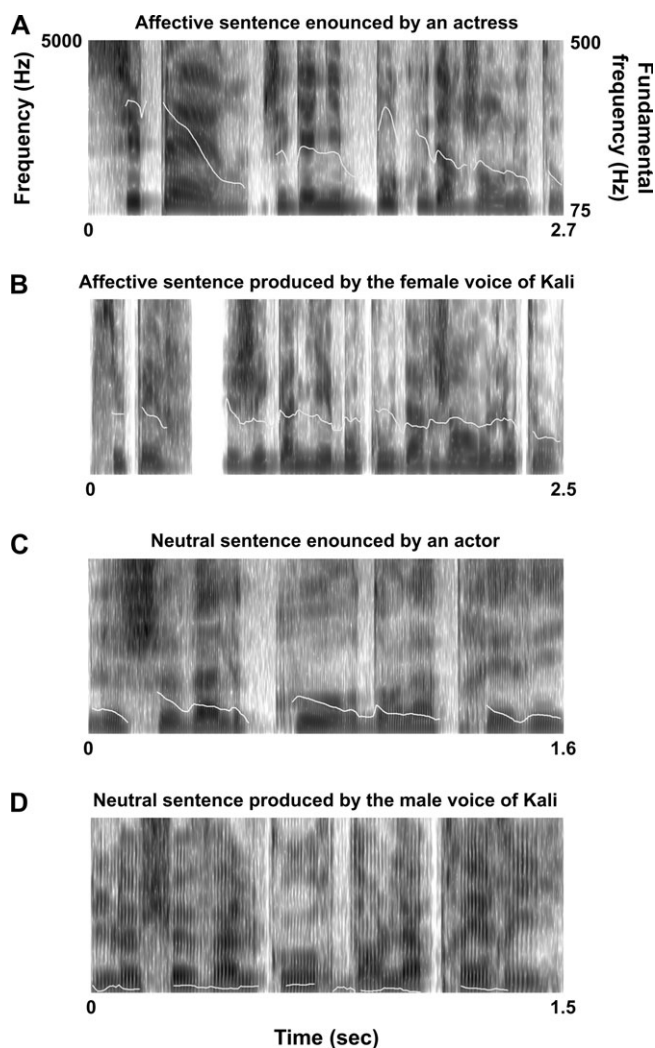


Figure 1. Spectrograms and pitch contours. Frequency (left ordinate) and fundamental frequency (white line, right ordinate) are given in function of time (abscissa, ms) for a sentence with affective semantic content (“Super, j’ai gagné beaucoup d’argent au loto”/“Great, I’ve won lot of money on the lotto”) (A) enounced by an actress and (B) produced by the female voice of Kali, a text-to-speech software, and for a sentence with neutral semantic content (“Le cheval court dans la prairie”/“The horse runs in the meadow”) (C) enounced by an actor and (D) produced by the male voice of Kali.

France). The session started with 2 anatomical acquisitions. First, a high-resolution structural T_1 -weighted sequence (T_1 -MRI) was acquired using a spoiled gradient-recalled sequence (SPGR-3D, matrix size = $256 \times 256 \times 128$, sampling = $0.94 \times 0.94 \times 1.5 \text{ mm}^3$) to provide detailed anatomic images and to define the location of the 32 axial slices to be acquired during both the second anatomical acquisition and the functional sequences. The second anatomical acquisition consisted of a double echo proton density/ T_2 -weighted sequence (PD-MRI/ T_2 -MRI, matrix size = $256 \times 256 \times 32$, sampling = $0.94 \times 0.94 \times 3.8 \text{ mm}^3$).

Each of the 8 functional runs consisted of a time series of 66 echo-planar T_2^* -weighted volumes (blood oxygen level-dependent [BOLD], time repetition = 6 s, echo time = 60 ms, anisotropic fraction = 90° , sampling = $3.75 \times 3.75 \times 3.8 \text{ mm}^3$). To ensure the signal stabilization, the first 3 BOLD volumes were discarded at the beginning of each run.

Preprocessing of Functional Images

The preprocessing was built on the basis of SPM99b subroutines ([Friston and others 1995; Ashburner and Friston 1999], AIR5.0 [Woods and others 1992], Atomia [Verard and others 1997]) locally developed, and encapsulated in a semiautomatic processing pipeline. The prepro-

cessing included 9 steps: 1) correction for differences in BOLD image acquisition time between slices; 2) rigid spatial registration of each of the BOLD volumes onto the fourth BOLD volume of the first acquired run (BOLD4); 3) computation of the spatial rigid registration and resampling matrices from BOLD4 to T_2 -MRI and PD-MRI to T_1 -MRI; 4) computation of the nonlinear registration matrix for stereotaxic normalization of the T_1 -MRI on the Montreal Neurological Institute T_1 -weighted templates (T_1 -MNI) (Collins and others 1994) (SPM99b stereotaxic normalization with 12-parameter rigid body transformations, $7 \times 8 \times 7$ nonlinear basis functions, 12 nonlinear iterations, medium regularization, bounding box in between -90 to $+91$ mm left-right, -126 to 91 mm back-front, and -72 to 109 mm feet-head directions, sampling $2 \times 2 \times 2 \text{ mm}^3$); 5) combination of the matrices computed at the previous 2 steps, visual checking and optional optimization of the EPI4 (echo-planar imaging 4) to T_1 -MNI registration in the stereotaxic space; 6) spatial resampling of each BOLD volume into the T_1 -MNI stereotaxic space; 7) spatial smoothing of each BOLD volume by a Gaussian filter (full width half minimum = $8 \times 8 \times 8 \text{ mm}^3$); 8) high-pass filtering (cut-off of 0.0102 Hz) of each voxel time course; and 9) normalization of the voxel values by the average of its value in the course of the 2 runs (i.e., across time course).

Statistical Analysis of Functional Images

The functional data were analyzed and integrated in a statistical model by the semiautomatic software SPM99b (Wellcome Department of Cognitive Neurology, www.fil.ion.ucl.ac.uk/spm/).

The individual data consisted of 8 contrast maps that presenting a BOLD signal increase covarying with the cognitive task compared with the control task (beep detection). These 8 contrast maps corresponding to 2 runs of grammatical classification on sentences produced by Kali, 2 runs of grammatical classification on sentences enounced by the actors, 2 runs of affective classification, 1 on emotional and 1 on attitudinal sentences produced by Kali, and 2 runs of affective classification, 1 on emotional and 1 on attitudinal sentences enounced by actors. Then a second-level analysis was performed including, for each subject, the 8 BOLD contrast maps. Because no significant difference was observed between the attitudinal and emotional runs at 0.05 corrected threshold for multiple comparisons, these 8 runs were collapsed into 4 contrast maps in a second-level analysis: one corresponding to the mean of emotional and attitudinal classification of sentences enounced by actors minus beep detection (AffAct); one corresponding to the mean of emotional and attitudinal classification of sentences produced by Kali minus beep detection (AffKali); one corresponding to the mean of both grammatical classification tasks performed on sentences enounced by actors minus beep detection (GrAct); and one corresponding to the mean of grammatical classification tasks performed on sentences produced by Kali minus beep detection (GrKali). In the second-level analysis, the following contrasts were computed:

1. (GrAct-GrKali) and (GrKali-GrAct) to evaluate the effect of the kind of speaker ($P \leq 0.001$ uncorrected threshold).
2. [(GrAct) \cap (GrKali)]: conjunction analysis of the grammatical classifications to evidence neutral sentence comprehension areas (0.0025 corrected thresholds for multiple comparisons, corresponding to 0.05 per contrast).
3. [(AffAct-GrAct) \cap (AffKali-GrKali)]: conjunction of the “affective minus grammatical classification” contrasts obtained when the sentences were produced by Kali (AffKali-GrKali) and enounced by actors (AffAct-GrAct) to evidence the areas involved in affective sentence comprehension (0.0025 corrected threshold, corresponding to 0.05 per contrast).
4. [AffAct-AffKali] to uncover areas dedicated to affective prosody processing ($P \leq 0.001$ uncorrected).

Hemispheric asymmetries of these networks were evaluated, thanks to a whole-brain approach. First, we computed asymmetrical contrast maps resulting from the subtraction of individual flipped contrast maps in their x axis with their corresponding nonflipped maps. This resulted in a map per subject and per condition corresponding to the difference between left and right BOLD value in each voxel of the left side of the maps and right minus left BOLD value on the right side of the maps.

Then a second-level analysis was performed on these asymmetrical contrast maps with the same design as the one we used for BOLD variations contrast maps. The asymmetries during AffAct were also investigated ($P \leq 0.001$ uncorrected).

Lastly, the BOLD variations for the local maxima detected as significant in a given contrast were plotted for every task to further characterize their activation profile in the 4 conditions.

Results

Behavioral Data

The ANOVA evidenced a significant Task \times Voice interaction concerning both the number of CA ($F_{19} = 89.6$, $P < 0.0001$, Fig. 2A) and the RT ($F_{19} = 19.9$, $P = 0.0006$, Fig. 2B).

This interaction was related to the fact that classification of affective sentences (affective classification) enounced by actors with affective prosody (AffAct) was performed faster and with greater accuracy than when this task was performed on sentences produced by Kali devoid of affective prosody (AffKali) (RT: AffAct = 447 ± 71 ms, AffKali = 510 ± 66 ms, $t_{19} = 5.3$, $P < 0.0001$; CA: AffAct = $84 \pm 9\%$, AffKali = $69 \pm 9\%$, $t_{19} = -7.9$, $P < 0.0001$), whereas this was not the case with the grammatical classification task. In other words, affective classification was easiest to perform in the presence of affective prosody.

Indeed the Voice \times Task interaction also seated into the fact that such an effect of Voice was not found during the grammatical classification of neutral sentences (this task will

further be called the “grammatical classification”): no significant difference was found whether the sentences were enounced by actors (GrAct) or produced by Kali (GrKali) (CA: GrAct = $87 \pm 8\%$, GrKali = $86 \pm 6\%$, paired $t_{19} = -0.9$, $P = 0.36$; RT: GrAct = 385 ± 82 ms, GrKali = 384 ± 89 ms, $t_{19} = 0.4$, $P = 0.7$).

Note that a significant Task main effect was evidenced: for both types of Voice, better performances were achieved during grammatical classification than during affective classification in terms of CA ($F_{19} = 43.7$, $P < 0.0001$) and RT ($F_{19} = 50.4$, $P < 0.0001$). A significant main effect of Voice was also observed independently of the task: greater CA ($F_{19} = 31.7$, $P < 0.0001$) and faster responses were found (RT: $F_{19} = 12.9$, $P = 0.002$) when the sentences were enounced by actors than when they were produced by Kali.

These results appeared very coherent with subjects that reported in the postsession questionnaire that the grammatical classification was the easiest to perform and that the affective sentences uttered by actors were easier to classify than affective sentences produced by Kali. All subjects indicated that, during affective classification, they used intonation cues to assess the sentences’ affective content when it was present, whereas they relied on the affective verbal content of the sentences to complete the task when sentences missed affective prosody (produced by Kali). In addition, in the presence of affective prosody (sentences enounced by actors), 18 subjects out of 23 (78%) still used sentences’ verbal content in addition to intonation. Note that 11 subjects (48%) happened to rehearse the sentences, whatever the speaker, to complete the affective classification.

Functional Imaging Results

Grammatical Classification of Neutral Sentences

The comparisons, either (GrKali-GrAct) or (GrAct-GrKali), detected no impact specific to the text-to-speech software or the actors on the cerebral network involved during grammatical classification (even when lowering the threshold to 0.001 uncorrected for multiple comparisons).

The conjunction analysis of the grammatical classification tasks performed on sentences uttered by actors and Kali [(GrAct-beep detection) \cap (GrKali-beep detection)] revealed massive leftward activations in the temporal, frontal, and parietal lobes (Fig. 3, Table 1). In the left temporal lobe, activations were identified in the superior temporal sulcus (STS) and superior temporal gyrus (STG), extending to Heschl’s gyrus, the planum temporale, and to the posterior part of the middle temporal gyrus. In the frontal lobes, the inferior frontal gyrus (IFG), the precentral gyrus, and the supplementary motor area (SMA) were activated. This network also included the parietal lobe stretching from the postcentral to the superior parietal gyrus. The calcarine sulcus, putamen, thalami, and cerebellar cortex also showed BOLD signal increase.

Although BOLD analysis evidenced mirror activations in the right hemisphere, the direct comparison of left and right activations (thanks to the asymmetrical contrast maps) showed a significant leftward lateralization of the activated areas, except for the cerebellar cortex, which instead was asymmetrical to the right.

Neural Substrate of Affective Sentence Comprehension

To identify the network implicated in affective sentence comprehension independently of the presence of affective

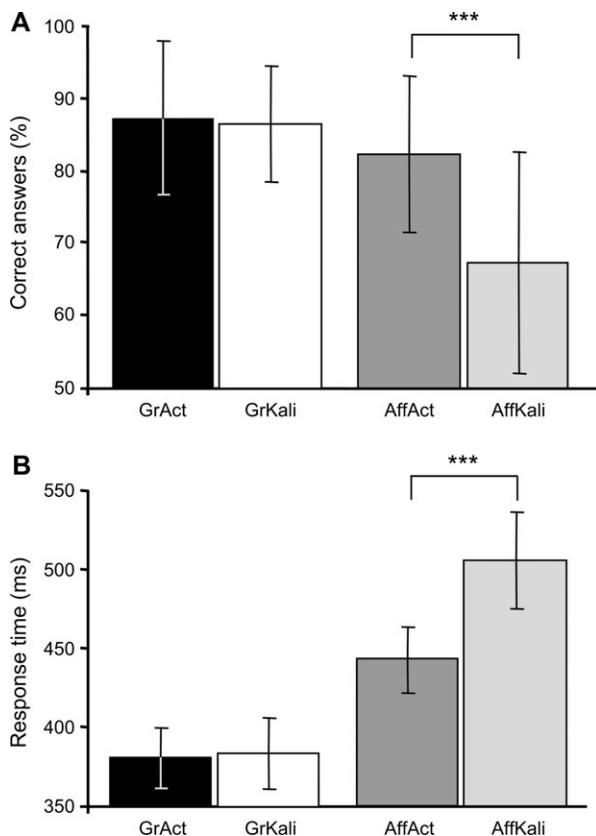


Figure 2. Behavioral results. (A) Average percentage of CA and (B) mean RTs (\pm standard deviation, ms) during affective classification of sentences with affective semantic content enounced by actors (AffAct, dark gray bar) or produced by Kali (AffKali, light gray bar), during the grammatical classification of sentences with neutral semantic content spoken by actors (GrAct, black bar) or produced by Kali (GrKali, white bar, *** $P \leq 0.001$).

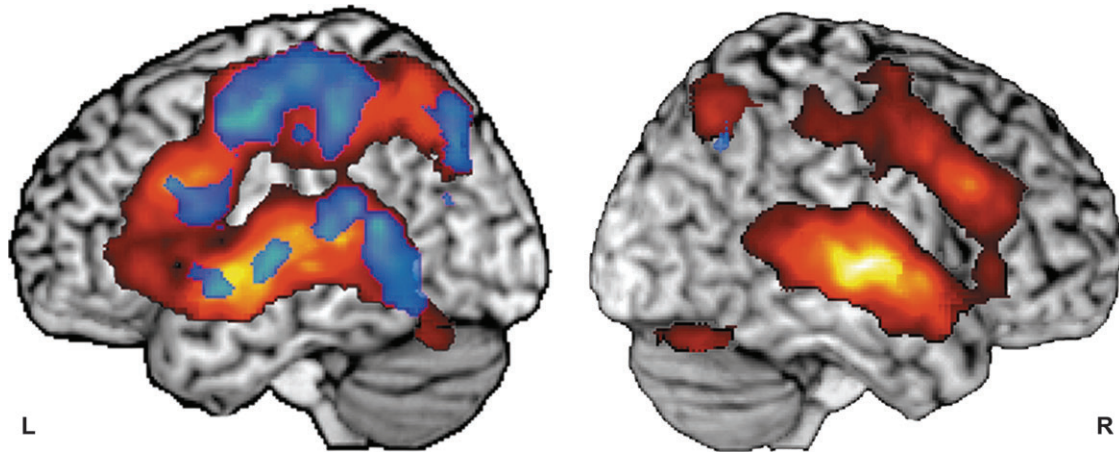


Figure 3. Cortical network engaged during the grammatical classification of sentences with neutral verbal content regardless of speaker. Cortical areas significantly activated during the grammatical classification of sentences produced either by Kali or by actors compared with beep detection are projected on the left (L) and right (R) hemisphere of MNI reference brain (conjunction analysis given at 0.0025 corrected threshold for multiple comparisons). Red scale is for BOLD signal variation; blue scale is for significantly asymmetrical BOLD variations in each hemisphere.

Table 1
Cortical areas implicated during grammatical classification of neutral sentences

Anatomical localization	N voxels	x	y	z	Z value	P corrected
Right hemisphere						
STG	12,316	64	-14	0	∞	<0.001
Middle temporal gyrus		70	-32	12	∞	<0.001
Inferior temporal gyrus		44	-68	-28	∞	<0.001
Precentral gyrus		52	20	28	∞	<0.001
IFG tri		58	32	20	∞	<0.001
Anterior insula		38	30	-2	7.20	<0.001
Superior parietal gyrus		34	-64	50	∞	<0.001
Postcentral gyrus		48	-28	42	7.82	<0.001
Calcarine		12	-68	4	∞	<0.001
Putamen		22	4	10	∞	<0.001
Cerebellar cortex ^a		30	-64	-26	∞	<0.001
Left hemisphere						
Precentral gyrus	28,192	-66	-12	4	∞	<0.001
Precentral gyrus ^a		-44	12	28	∞	<0.001
Precentral gyrus ^a		-54	-8	50	∞	<0.001
IFG tri		-48	-24	10	∞	<0.001
IFG tri ^a		-58	18	16	∞	<0.001
SMA		-2	0	64	∞	<0.001
Cingulate gyrus		10	14	48	∞	<0.001
Anterior insula		-34	28	0	∞	<0.001
STG ^a		-52	-16	2	∞	<0.001
STG ^a		-60	-50	14	∞	<0.001
Heschl gyrus		-40	-32	14	∞	<0.001
Middle temporal gyrus		-64	-26	0	∞	<0.001
Inferior temporal gyrus		-48	-60	-16	∞	<0.001
Postcentral gyrus ^a		-46	-38	50	∞	<0.001
Superior parietal gyrus		-30	-62	46	∞	<0.001
Precuneus		20	-74	-24	∞	<0.001
Putamen ^a		-22	6	8	∞	<0.001
Calcarine		-8	-86	-2	∞	<0.001
Thalamus		-12	-16	8	∞	<0.001
Cerebellar cortex		-30	-64	-26	∞	<0.001

Note: Stereotaxic coordinates are given for the significant clusters issued from the conjunction of the grammatical classification of sentences enounced by actors and produced by Kali. Tri, pars triangularis.

^aAreas exhibiting significantly larger BOLD variations than their contralateral counterparts. Both analyses were conducted at $P \leq 0.0025$ corrected for multiple comparisons, corresponding to a 0.05 threshold for each contrast.

prosody, we computed a conjunction analysis of the differences between the affective and the grammatical classification obtained when the sentences were spoken by actors and when the sentences were produced by Kali [(AffAct-GrAct) \cap (AffKali-GrKali)].

The clusters, where greater activity was observed during affective classification than during grammatical classification, could be split according to the profile of their BOLD signal variation calculated at the local maximal peak of activity as the mean BOLD values in each condition (Fig. 4, Table 2).

A first set of areas was composed of the anterior and inferior part of the bilateral IFG, the bilateral anterior insula, the pre-SMA ($y > 26$ mm), the subcortical areas (left thalamus and right caudate nucleus), and the right cerebellar cortex. These regions already activated by the grammatical classification further showed increased activity during affective classification whether the sentences included affective prosody or not.

The second set of areas located in the medial superior frontal gyrus (MF1) and at the left posterior ending of the STS (pSTS) was activated during affective classification, but they presented a negative BOLD signal variation during grammatical classification.

Comparison of the left and right hemisphere activations in the contrast [(AffAct-GrAct) \cap (AffKali-GrKali)] using a whole-brain approach (thanks to the asymmetrical contrast maps) evidenced a significant leftward asymmetry in the pars triangularis/orbitalis of the IFG and in the pSTS, provided that no activity was detected in the right pSTS.

One should note that the reverse comparison (grammatical minus affective classification) did not reveal any differences at 0.001 uncorrected threshold for multiple comparisons.

Cerebral Network for the Processing of Affective Prosody

The areas involved in affective prosody processing were uncovered in the difference between brain activity during affective classification of sentences enounced by actors and during affective classification of sentences produced by Kali lacking affective prosody (AffAct-AffKali).

An activation located in the right anterior part of the STS (aSTS) passed the corrected threshold (0.05) (Fig. 5). When the threshold was lowered (0.001 uncorrected for multiple comparisons), this temporal activation spread to the bilateral anterior part of STG, including Heschl's gyri and to the right posterior part of STG (pSTG). At this threshold, the bilateral amygdalae, the putamen, and the hippocampal gyri showed a higher activity when affective prosody was present, as did

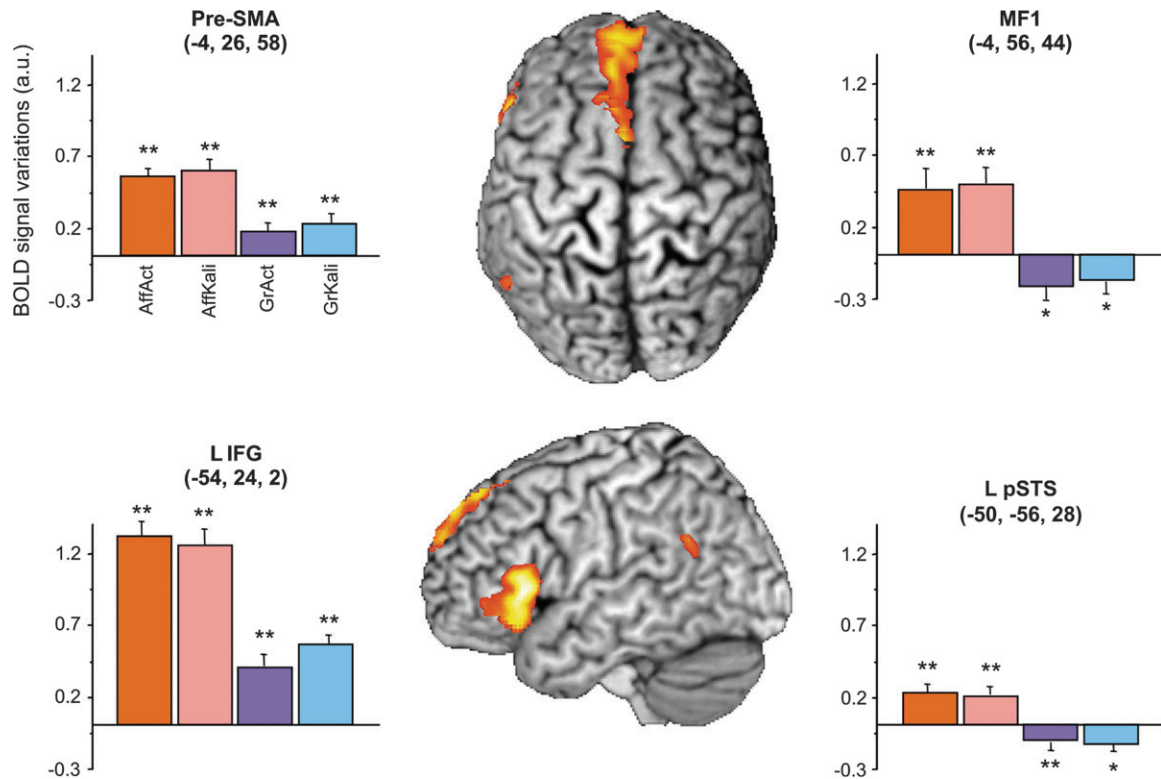


Figure 4. Brain areas more activated by affective than grammatical classification. Conjunction analysis of the affective minus grammatical classification of sentences enounced by actors (respectively, AffAct and GrAct) with the same contrast on sentences produced by Kali (respectively, AffKali and GrKali) overlaid on MNI-referenced brain template ($P \leq 0.0025$ corrected threshold for multiple comparisons). Areas evidenced by this conjunction and thus showing a significant larger activity when the sentences included an affective content whatever the presence of affective prosody than during grammatical classification of neutral sentences were located in the pre-SMA, the MF1, the left IFG (L IFG), and the left pSTS (L pSTS). Bar charts provide the average BOLD signal variation during each condition compared with the beep detection reference task at the local maximal peak of activity (error bars correspond to standard error of the mean; peaks coordinates are given in stereotaxic coordinates in mm; * $P < 0.05$, ** $P < 0.01$ correspond to the results of one sample t -tests comparing the BOLD signal variation during each conditions with beep detection reference task; a.u., arbitrary unity; AffAct, red bar; AffKali, pink bar; GrAct, blue bar; GrKali, purple bar).

Table 2

Cortical areas showing higher activity during affective classification of sentences containing or not affective prosody than during grammatical classification

Anatomical localization	<i>N</i> voxels	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z</i> value	<i>P</i> corrected
Frontal lobe						
L IFG tri ^a	777	-54	24	2	∞	<0.001
L IFG orb ^a		-46	42	-2	6.28	<0.001
L Anterior insula	59	-30	28	0	6.40	<0.001
L Medial part of F1	1,270	-4	56	44	∞	<0.001
L Medial part of F1		-6	44	52	7.72	<0.001
L Medial part of F1		-4	60	28	7.71	<0.001
R IFG tri	66	58	26	-2	6.79	<0.001
R IFG orb		52	38	-12	4.80	<0.001
Temporal lobe						
L STS ^a	108	-50	-56	28	6.74	<0.001
Subcortical regions and cerebellum						
L Thalamus	253	-4	-10	10	6.85	<0.001
R Caudate nucleus ^a	79	12	4	12	6.51	<0.001
R Cerebellar cortex	111	26	-78	-28	∞	<0.001
R Cerebellar cortex		12	-82	-26	7.06	<0.001
L Cerebellar vermis	105	-4	-60	-36	7.05	<0.001
R Cerebellar vermis		4	-60	-34	6.78	<0.001
L Cerebellar cortex	26	-6	-82	-22	6.44	<0.001

Note: Stereotaxic coordinates are given for the significant clusters of the conjunction analysis of the affective minus grammatical classification of sentences enounced by actors and produced by Kali ($P \leq 0.0025$ corrected for multiple comparisons). F1, superior frontal gyrus; orb, pars orbitaris; L, Left; R, Right.

^aAreas exhibiting significantly larger BOLD variations than their contralateral counterparts ($P \leq 0.05$ corrected threshold).

motor areas, namely, the bilateral precentral gyri and right SMA (Table 3).

Analysis of the mean BOLD signal values calculated for each of the 4 conditions in the local maximal peak of each cluster demonstrated that these areas presented different profiles. The right aSTS and pSTG activated during grammatical classification showed a further increase in activity when the sentences included both an affective semantic content and affective prosody (Fig. 5). On the other hand, the frontal regions, the left temporal areas, and the amygdalae mainly exhibited a reduction or no increase in activity during the affective classification of affective sentences lacking affective prosody (Fig. 6).

Hemispheric Lateralization of Temporal Areas Recruited by Affective Prosody

As stated in the Introduction, a key issue raised by previous neuropsychological and functional imaging literature concerns the right hemisphere dominance of temporal areas for prosodic processing. Based on the a priori hypothesis that prosodic temporal areas should exhibit a rightward lateralization during prosodic processing, we investigated the significant asymmetries in the AffAct contrast. Only considering the temporal lobe, significant rightward asymmetries were present in a lateral subpart of aSTS ($x = 64$, $y = -6$, $z = -12$, Z score = 4.30, extent = 29 voxels) and in an internal subpart of pSTG ($x = 50$, $y = -38$, $z = 6$,

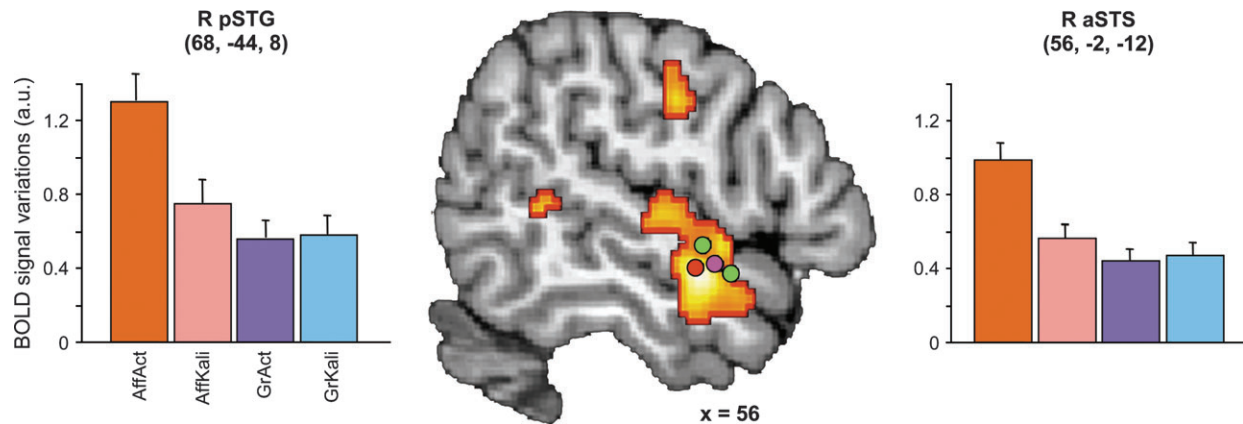


Figure 5. The right temporal areas and affective prosody. The right aSTS areas (R aSTS) and the right pSTG (R pSTG) that were more activated during affective classification in the presence of affective prosody (sentences enounced by actors) than in the absence of affective prosody (sentences produced by Kali) are superimposed on a sagittal slice of the MNI-referenced brain ($x = 56$, $P \leq 0.001$ uncorrected threshold). The cluster R aSTS, which shows a greater BOLD signal increase in presence of affective prosody than in other conditions, is located close to HSVA as defined in ● Belin and others (2000), ● Belin and Zatorre (2003), and ● Kriegstein and others (2003). Bar charts provide the average BOLD signal variation during each condition compared with the beep detection reference task at the local maximal peak of activity (error bars correspond to standard error of the mean; peak coordinates are given in stereotaxic coordinates in mm; a.u., arbitrary unity; AffAct, red bar; AffKali, pink bar; GrAct, blue bar; GrKali, purple bar).

Table 3

Cortical areas implicated in affective prosody comprehension

Anatomical localization	N voxels	x	y	z	Z value	P uncorrected
R STS/STG	552	56	-2	-12	4.98	<0.001
R STS/STG		60	4	-4	3.82	<0.001
R Heschl gyrus		54	-12	4	3.65	<0.001
R Amygdala	292	30	-2	-12	4.60	<0.001
R Hippocampus		32	-14	-2	3.77	<0.001
R Putamen		26	4	-2	3.64	<0.001
R STS	268	68	-44	8	4.32	<0.001
R STS		70	-44	16	3.91	<0.001
R STS		60	-46	4	3.74	<0.001
R SMA	89	6	-22	66	3.99	<0.001
L Postcentral gyrus	136	-62	-10	42	3.98	<0.001
L Putamen/amygdala	59	-28	-4	-12	3.86	<0.001
R Central sulcus	82	54	-8	34	3.85	<0.001
L Middle insula	148	-44	2	0	3.83	<0.001
L STG		-56	2	0	3.68	<0.001
L Central sulcus		-56	-8	30	3.60	<0.001
L Posterior insula	122	-42	-14	10	3.77	<0.001
L Heschl gyrus		-36	-22	4	3.57	<0.001
L Putamen		-30	-14	4	3.29	0.001
R Posterior insula	37	38	-20	18	3.74	<0.001

Note: Stereotaxic coordinates of clusters obtained in the contrast affective classification in presence of affective prosody (sentences were enounced by actors) minus affective classification in absence of affective prosody (sentences were produced by Kali, $P \leq 0.001$ uncorrected for multiple comparisons). L, Left; R, Right.

Z score = 4.14, extent = 101 voxels). To provide a detailed description of the behavior of these areas that were detected as asymmetrical, we calculated individually the BOLD signal variations in these clusters on each side (in contrast maps and flipped maps) and for each condition.

A repeated-measures ANOVA was then performed on these clusters BOLD values, entering Side (right vs. left hemisphere) and Speaker (Kali vs. Actor) as factors. During affective classification, a significant interaction between Side and Speaker, that is, effect of the presence of affective prosody, was observed (Fig. 7; aSTS: $F = 5.6$, $P < 0.05$; pSTG: $F = 8.4$, $P < 0.01$). This interaction was related to a larger BOLD increase in right than left areas when affective prosody was present. A main effect of hemisphere was observed, confirming the larger involvement of the right temporal areas during the affective classification whether affective prosody was present or not (aSTS: $F = 4.7$,

$P < 0.05$; pSTG: $F = 6.1$, $P < 0.05$). A main effect of affective prosody was also found, showing that temporal areas were more involved when sentences included affective prosody than when they lacked it (aSTS: $F = 39.7$, $P < 0.0001$; pSTG: $F = 72.8$, $P < 0.0001$). Note that during the grammatical classification, the main effect of neither Side (aSTS: $F = 0.7$, $P > 0.05$; pSTG: $F = 2.1$, $P > 0.05$) nor Speaker (aSTS: $F = 0.8$, $P > 0.05$; pSTG: $F = 0.8$, $P > 0.05$) was observed.

Discussion

The present paradigm allowed to disentangle the areas involved in affective prosody from those involved in affective semantic and syntactic processing during affective sentence comprehension. First, the use of a reference condition involving the comprehension of sentences with a neutral emotional content allowed to pull out the areas dedicated to affective discourse comprehension independently of the presence of affective prosody. These areas were the left inferior frontal area, pSTS, and MF1. Interestingly, whereas IFG was already engaged during grammatical classification of neutral sentences, pSTS and MF1 were specifically involved when an emotional verbal material was present. Second, the use of a text-to-speech software that included grammatical but not affective prosody allowed to uncover areas involved in prosodic processing, in conditions with equivalent semantic content. These areas were located within the right temporal lobe and presented a rightward asymmetry, as could have been expected from studies of aprosodia. They corresponded to the human-selective voice area (HSVA) and the integrative posterior temporal cortex.

Network for the Grammatical Classification of Neutral Sentences

Although Kali, the text-to-speech software, sounded natural, we needed to check for its possible impact on neutral sentence comprehension. During the grammatical classification, we observed no differential effect on behavioral results of sentences with neutral verbal content produced by Kali compared with those uttered by actors, demonstrating the good intelligibility and correct grammatical prosody of this software

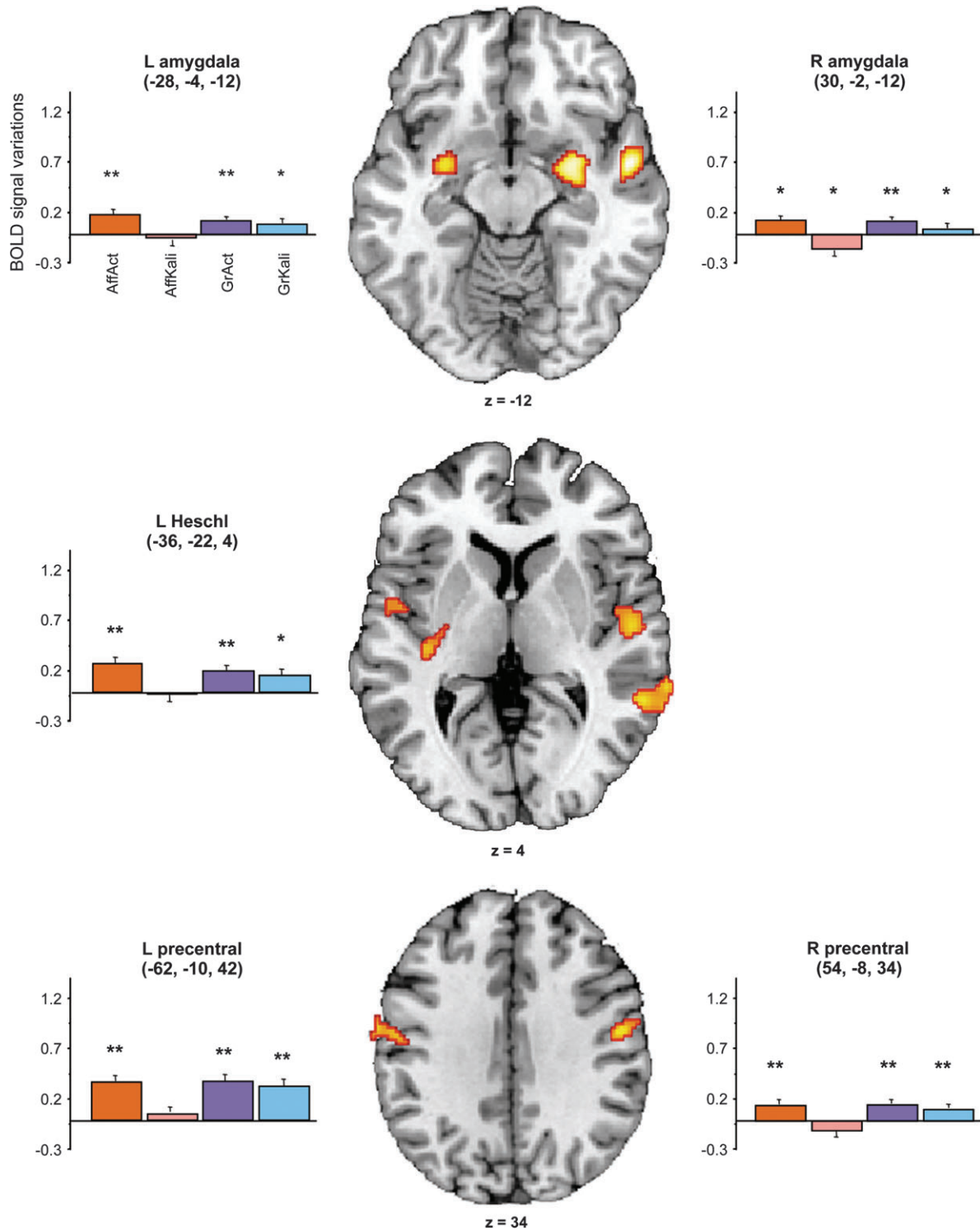


Figure 6. Areas showing decreased activity when affective semantic sentences lacked affective prosody. The clusters in bilateral amygdala, left heschl gyrus, and bilateral precentral gyrus, which were obtained in the contrast of affective classification in the presence of affective prosody versus in the absence of affective prosody are represented on axial slices of the MNI brain. Bar charts provide the average BOLD signal variation during each condition compared with the beep detection reference task at the local maximal peak of activity (error bars correspond to standard error of the mean; peaks coordinates are given in stereotaxic coordinates in mm; * $P < 0.05$, ** $P < 0.01$ correspond to the results of one sample t -tests comparing the BOLD signal variation during each conditions with beep detection reference task; a.u., arbitrary unity; AffAct, red bar; AffKali, pink bar; GrAct, blue bar; GrKali, purple bar).

compared with the natural stimuli. In the same vein, functional results did not show any difference during grammatical classification of sentences produced by Kali or enounced by actors. This is very likely because Kali built the speech stimuli from

a database of naturally spoken syllables. One study on the impact of synthetic speech on neural activity found greater activity in the left premotor cortex during listening to natural speech than during listening to synthetic speech (Benson and others 2001),

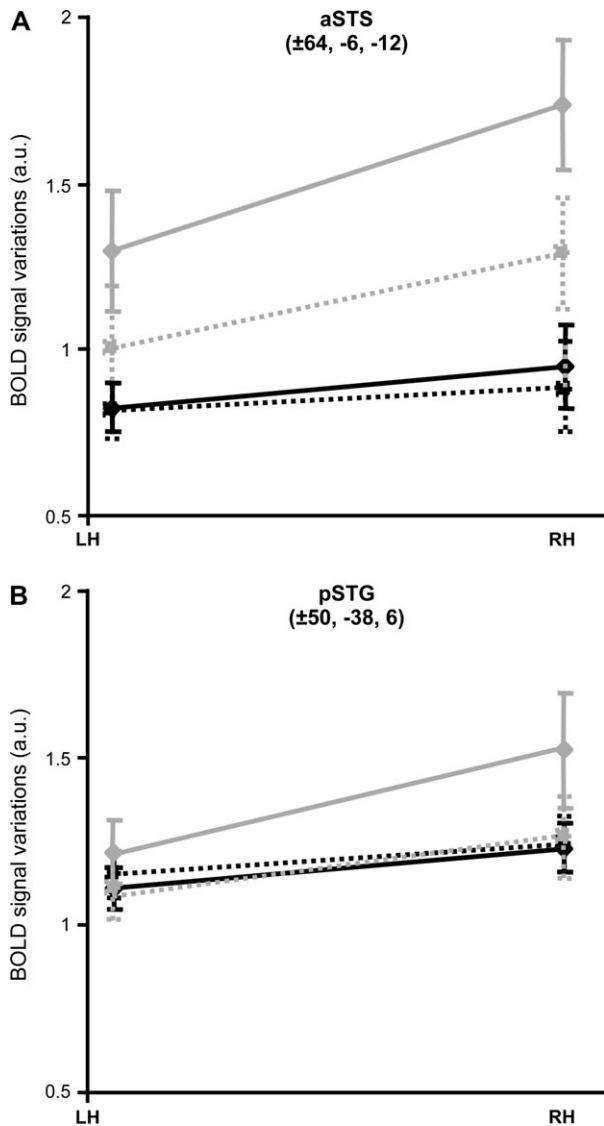


Figure 7. Lateralization of temporal areas in presence of affective prosody. Variation of the BOLD signal during the 4 conditions in the clusters corresponding to the local maximal peaks in (A) the aSTS and (B) the pSTG (—◆— AffAct, —■— gray line; AffKali, —●— gray dotted line; GrAct, —■— black line; GrKali, dark dotted line; LH, left hemisphere; RH, right hemisphere).

but these authors used synthetic stimuli that were not composed of natural tokens.

The grammatical classification of neutral sentences thus appears to be a relevant reference task to remove from the affective classification neural network: 1) IFG and STG activity related to sentence processing (Vigneau and others 2006), 2) right frontoparietal network engagement for attention, anticipation, and selection of the response (Tzourio and others 1997), and 3) activation of pre- and postcentral gyri that corresponded to the sensory-motor cortical representation of the hand (Mesulam 2000) activated by the motor response.

Network for Affective Semantic Comprehension

Semantic and Emotional Frontal Areas

A frontal network was recruited during affective classification of sentences with affective semantic content and to a lesser extent

during grammatical classification of neutral sentences. Although homologous rightward activity was present, the significant leftward asymmetry of this network attested its language specificity. These areas were located in the anterior and inferior part of the left IFG, known to be involved in semantic categorization (Poldrack and others 1999; Adams and Janata 2002) and selection of semantic knowledge (Wagner and others 2001; Booth and others 2002). They were easy to relate to the strategy, reported by all subjects, of relying on semantic cues to classify sentences with affective verbal content. In addition, subjects reported to mentally rehearse the sentences, a strategy most likely corresponding to the observed activations in pre-SMA and the left anterior insula known to be involved in speech mental articulation (Ackermann and Riecker 2004).

The present IFG clusters located in the pars orbitalis overlapped the areas activated during the emotional discrimination of sentences compared with the repetition of the last word of these sentences (George and others 1996). They also overlapped in studies comparing the judgment of emotional prosodic accentuation (Wildgruber and others 2004) or comparing emotional discrimination with verbalization of a target vowel (Wildgruber and others 2005). This orbitofrontal area is also activated in the current work by the processing of sentences with affective verbal content independently of the presence of affective prosody, in line with earlier suggestions of the role of this region in emotional processing (Wildgruber and others 2004, 2005). It also agrees with the report of activation of the pars orbitalis during the perception of emotional words (Beaugard and others 1997) or gender discrimination operating on an emotional face (Blair and others 1999).

Medial Prefrontal and Left pSTS Activations: Inference of the Speaker's Mental State

A second set of regions, namely, the MF1 and the left pSTS, showed increased activity when subjects performed the affective classification whereas they were not activated during the grammatical classification.

Involvement of the medial wall of the frontal lobe could be related to error detection (Botvinick and others 2004). As a matter of fact in the present study, behavioral results showed a larger number of errors during the affective classification than during the grammatical classification, a difference that could be related to this higher MF1 activity during the affective classification task. However, this hypothesis is challenged by the numerous reports that located the region sensitive to error detection in the anterior part of the cingulate gyrus, in a lower location than the cluster of the present study (for reviews, see Bush and others 2000; Ridderinkhof and others 2004; Rushworth and others 2004).

Actually, numerous studies on theory of mind (TOM) processing intersected in MF1 activation found in the present study, as shown in Figure 8A (for methods, see Jobard and others 2003). The expression TOM refers to the ability to explain and predict one's own actions and those of other intelligent agents (Premack and Woodruff 1978). The tasks used in these previous TOM studies involved either verbal (Vogel and others 2001; Harris and others 2005) or visual material (films [Castelli and others 2000], cartoons [Brunet and others 2000; Gallagher and others 2000; Walter and others 2004], or objects [Goel and others 1995]) and included the inference of another's mental state, such as the attribution of intention

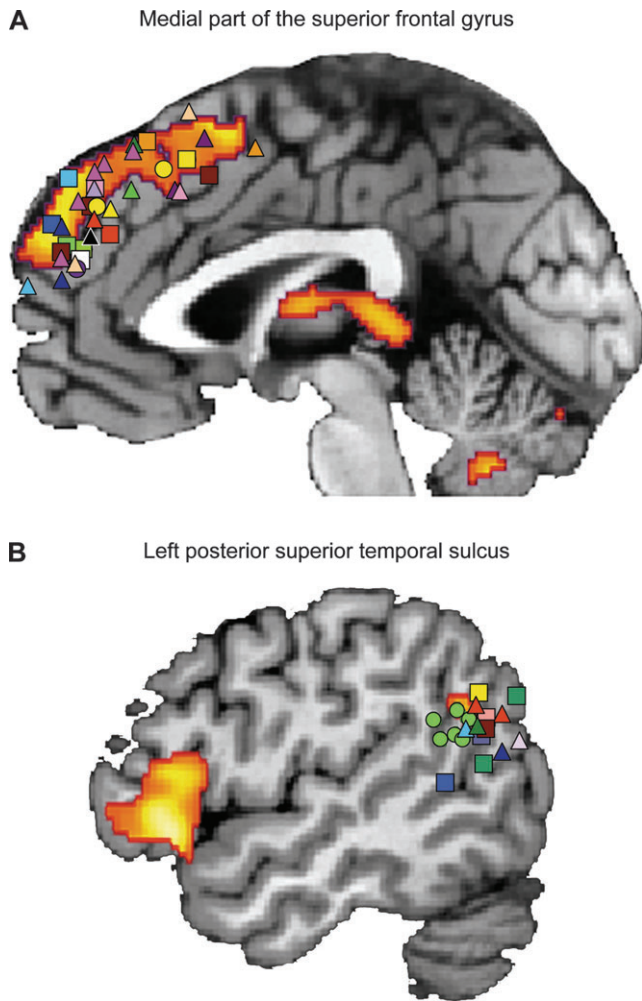


Figure 8. Meta-analysis in the medial frontal gyrus and pSTS: activations related to the processing of affective sentences are superimposed on the internal surface and sagittal slice of MNI single subject. Peaks issued from studies dealing with TOM (squares), self (triangles), emotion (pink and purple circles), and syntactic processing (green circles) are represented. (A) Projection on the medial surface and (B) on the sagittal slice ($x = -50$) of the MNI-referenced brain template of 1) activation detected with a conjunction analysis of the affective minus grammatical classification of sentences enounced by actors and produced by Kali (from red to yellow, $P \leq 0.0025$ corrected threshold for multiple comparisons); peaks of activation coming from 2) studies on **TOM processing** such as judging intentionality (Brunet and others 2000; Castelli and others 2000; Iacoboni and others 2004; Walter and others 2004; Harris and others 2005), comprehension of TOM stories (Fletcher and others 1995; Gallagher and others 2000; Vogeley and others 2001; Ferstl and Von Cramon 2002; Saxe and Kanwisher 2003), and judging other knowledge (Goel and others 1995; Ruby 2004), 3) studies on **self-reflection** such as self preference's judgment (Craig and others 1999; Sugiura and others 2004; Johnson and others 2005; Ochsner and others 2005; Schmitz and Johnson 2005), self knowledge's evaluation (Ruby and Decety 2003; den Ouden and others 2005), empathic situations (Lawrence and others 2006; Mitchell, Banaji, and Macrae 2005; Hynes and others 2006; Vollm and others 2006), self-evaluation of emotional content (Reiman and others 1997; Ochsner and others 2002), 4) studies on **emotional processing**: comparing the processing of emotional with neutral words (Beaugard and others 1997) or faces (Blair and others 1999), or 5) studies on the **integration** of semantic and syntactic processing at the level of sentences (Embick and others 2000; Kuperberg and others 2000; Kircher and others 2001; Luke and others 2002) or texts (Goel and others 1998; Homae and others 2002). All peaks of activations were placed in the MNI stereotaxic space (for methods, see Jobard and others 2003).

(Castelli and others 2000; Walter and others 2004; Harris and others 2005) and the observation of social interactions (Iacoboni and others 2004). It is also involved when one has to evaluate his/her own mental state (Craig and others 1999; Ruby and Decety 2003; Sugiura and others 2004; den Ouden and others 2005; Johnson and others 2005; Ochsner and others 2005; Schmitz and Johnson 2005) or his/her own emotional state (Reiman and others 1997; Ochsner and others 2002) (Fig. 8A). Emotional content of the stimuli appears crucial because MF1 is activated by the perception of empathic situations, when one had to infer and share the emotional experiences of others (Lawrence and others 2006; Mitchell, Banaji, and Macrae 2005; Mitchell and others 2005a, 2005b; Hynes and others 2006; Vollm and others 2006). Upper MF1 involvement during other emotional processing than empathy is seldom: in a review conducted by Phan and others (2002), it is the lower part of MF1 that is targeted by emotional processes, only few peaks of activation elicited by the perception of facial emotions (Blair and others 1999), or emotional words (Beaugard) overlapped with the part of MF1 activated in the present study (Fig. 8A). Thus, the upper MF1, activated during affective classification, is very likely involved in the representation of internal mental states—whether it is self-reflection (Northoff and Bermphohl 2004) or other's mental state that has to be inferred (Gallagher and Frith 2003)—a neural activity that appears to be enhanced by the emotional content of the stimuli to process (Gallagher and Frith 2003).

Concerning the left pSTS area activation, it is likely to be related to the integration of semantic and syntactic processing, crucial to succeed affective sentences classification, whereas useless to perform the grammatical classification. As a matter of fact, together with IFG they constitute a network for semantic analysis (Vigneau and others 2006). As illustrated in Figure 8B, this leftward lateralized area overlaps with peaks elicited by sentence-processing tasks that necessitate a semantic integration: judgment on grammatical errors compared with pronunciation errors (Embick and others 2000), generation of the final word of a sentence, (Kircher and others 2001), and comprehension of coherent rather than incoherent sentences (Kuperberg and others 2000; Luke and others 2002). Such a role in the semantic integration of complex verbal material is not limited to sentences; this area is also involved during text comprehension, with increased activation when sentences constitute a dialog (Homae and others 2002) or compose a syllogism (Goel and others 1998) than when they are not linked.

Interestingly, one should note that this role of the left pSTS in text integration includes a specific involvement during the comprehension of TOM stories. If some authors have proven that TOM stories engaged left pSTS more than unlinked sentences, confirming its role in semantic integration of complex material (Fletcher and others 1995; Ferstl and Von Cramon 2002), others have demonstrated an additional increase in activity of this region when they compared TOM stories with syntactically correct stories that described non-TOM events (Gallagher and others 2000; Saxe and Kanwisher 2003). This region is also engaged when one has to interpret others' intentions (Castelli and others 2000; Walter and others 2004), as well as when the representation of the self is needed such as during self-evaluation (Ruby and Decety 2003; den Ouden and others 2005; Johnson and others 2005), the processing of empathic situations (Hynes and others 2006; Vollm and others

2006), or valence assessment of emotional film (Lane and others 1997; Reiman and others 1997).

These observations conduct us to hypothesize that the role of the left pSTS, in the present study, cannot be restricted to the processing of the sentences propositional content. We instead postulate that the left pSTS would integrate the semantic and emotional content of speech to interpret the intended meaning of the speaker. The fact that the left pSTS together with MF1 were described as part of the core system for TOM processing (Gallagher and Frith 2003) suggests that activation of this network could be related to the computation of the speaker's mental state during affective classification. But considering the fact that activations in the upper MF1 and pSTS were also elicited by tasks relying on the self-evaluation of feelings or emotions (Fig. 8), subjects may as well have based their evaluation on a reflection about their own emotional state.

Note that this involvement of the MF1 or pSTS areas was independent of the presence of affective prosody because there was no observable modification in their activity when a lack of affective prosody increased the difficulty of the affective classification. Subjects' performances were relatively accurate in this condition (70% CA) leading to the conclusion that the possible call for TOM processing in emotional speech comprehension would be triggered by the affective semantic message rather than by the affective prosodic content of the sentences.

The Role of the Right Temporal Areas in Affective Prosody Processing

Although they were not explicitly informed of the presence of affective prosody, all subjects reported to rely on intonation to solve the affective classification task in the presence of affective prosody. Their greater speed and accuracy during this task led to the conclusion that optimization of affective discourse comprehension by the presence of affective prosody is supported by 2 areas in the right temporal lobe.

The presence of affective prosody led to the activation of the right aSTS that closely matches the so-called HSVA (Belin and others 2004). The HSVA was defined as a bilateral region that responds more to the human vocal sounds than to the environmental sounds (Belin and others 2000; Kriegstein and others 2003) or to the vocalization of other species (e.g., monkeys) (Fecteau and others 2004b), and its activity increases even more when several speakers are heard (Belin and Zatorre 2003) (Fig. 5). In line with the present result, these findings imply the involvement of the right HSVA in the processing of affective prosody, a human-specific acoustical feature.

More precisely, the right HSVA is implicated in the treatment of the paralinguistic features of voice that allow identification of speaker's gender (Fecteau and others 2004a). This paralinguistic function was confirmed by Grandjean and others (2005), who identified a rightward asymmetry of HSVA when subjects had to detect the gender of the speaker during presentation of pseudosentences to the left ear. Based on the present results, we hypothesized that the right HSVA computed the emotional content of voice through the extraction of slow acoustical elements that characterized affective prosody. Indeed, this process had been previously evidenced as a right lateral temporal lobe expertise (Belin and others 1998; Griffiths and others 1998; Meyer and others 2002; Mitchell and others 2003; Wildgruber and others 2005).

The right pSTG was the second area that showed greater activity in presence of affective prosody in this study. This result recalls Ross's model on neural correlates of affective prosody: from observations of aprosodic patients, he postulated that the rightward cortical organization for affective prosodic comprehension parallels the leftward organization of propositional language (Ross 1981). Indeed, right pSTG can be considered as homologous to Wernicke's area. Thus, we hypothesize that the right pSTG would perform the first interpretation in terms of emotional labeling of the relevant prosodic features extracted in the right HSVA. This information would be further integrated with the linguistic information computed in the left homologue via transcallosal transfer in order to complete sentence comprehension (Ross and others 1997).

The significant rightward asymmetry in HSVA and pSTG observed in the present study allows reconciliation of both neuropsychological and functional views: it shows that affective prosody processing led to bilateral but rightward asymmetrical activation in the temporal areas essential for affective prosodic comprehension (Ross 1981). This finding reinforces the assumption that, in functional studies of affective prosody, additional leftward semantic resources were engaged to try to catch the meaning of filtered sentences (Meyer and others 2002; Kotz and others 2003) or sentences constructed with pseudo-words (Price and others 1996; Grandjean and others 2005).

Reduction of Activity in the Audio-Motor Loop and the Amygdalae when Prosody Is Incongruent with Semantic Affective Content

Like the temporal areas, the amygdalae, precentral gyri, and left Heschl's gyrus exhibited greater activation during affective classification performed on sentences containing affective prosody than when this task was performed on sentences spoken without affective prosody. But as opposed to temporal regions, they were identified because of a decrease in activity during the affective classification task in the absence of affective prosody, rather than because they were activated by the presence of affective prosody (same amount of activity as during grammatical classification). This decrease was not related to the use of Kali, which had no impact on the neural activity of these areas during the grammatical classification (Fig. 6). Rather, this decrease appeared related to the fact that when Kali produced affective sentences, their prosodic and affective verbal content were not congruent. As a matter of fact, these Kali-produced sentences contained only grammatical prosody that has a neutral valence on affective scaling incongruent with the sentences' strong affective semantic content. This decrease in activity can be interpreted as the suppression of the incongruent prosodic processing that interfered with the comprehension of affective sentences. Indeed, decreases in BOLD signal can be considered as indicators of reduced input and local computation in the cortical areas (Logothetis and others 2001).

In the present case, the suppression at work when the prosodic and affective messages were not congruent targeted 2 systems. First was the processing of the affective content carried by the voice in the amygdalae (for a review, Adolphs 2002). The amygdala appears to be involved in emotional processing of voice as Scott and coworkers identified a deficit of emotional prosodic comprehension following lesions of the amygdalae (Scott and others 1997). Its decrease in activity

would thus suggest the intervention of a filtering process that reduced the emotional processing of the inadequate prosody. The second system, made up of the left Heschl and precentral gyrus, composes the audio-motor loop described by Hickok and Poeppel (2000) that enters into the processing of speech comprehension through an audio-motor simulation (Liberman and Whalen 2000). In the present case, the simulation of the speech that includes discordant prosody is very likely to be attenuated, possibly to allow the subjects to generate a more adequate prosody through mental imagery (Pihan and others 1997), as has been reported by some of them.

General Conclusion

This study allowed to disentangle the networks involved in semantic and prosodic processing of emotional discourse. It reconciles views issued from neuropsychology of aprosodia and functional imaging reports by confirming that the right temporal lobe is essential for emotional prosody processing and presents a rightward lateralization. Indeed, it is the right HSVA, together with the pSTG that process the emotional prosody. In addition, the use of sentences with equivalent syntactic and semantic content allowed to demonstrate that the involvement of the pars orbitalis of the right IFG was not linked with the presence of emotional prosody per se but rather with the presence of emotional words.

The present results open a new perspective: specific to emotional discourse is the activation of systems leading to the configuration of brain activity toward human social interactions. First, this is visible in the identification of HSVA as the region that processes emotional prosody. As a matter of fact, the role of HSVA in the right hemisphere can be expanded to the processing of social interactions charged with emotion. Right-damaged patients not only present a deficit of affective prosody comprehension (Ross 1981) but also exhibit a joint impairment in identification of faces depicting emotion (Blonder and others 1991). Indeed, HSVA exhibits preferential connectivity with the right fusiform face area (FFA) during speaker identification (Kriegstein and others 2005), implying that the paralinguistic function of this right area can be extended not only to the voice processing but also to the face processing. Such a close interaction seems to be crucial during social interaction because autistic children, who are characterized by impaired social interaction, present deficits in the right HSVA (Gervais and others 2004) as well as in FFA (Pierce and others 2001; Schultz 2005). The second result that reinforces this hypothesis is the possible involvement of TOM processing during the understanding of emotional discourse, which goes along with the observation by Brüne (2005) that TOM is permanently “online” in humans, screening even nonliving objects for putative intentions. Although it remains to be demonstrated, we believe that it is the cooperation of TOM and emotional systems that allows adequate verbal human communication. As a matter of fact, schizophrenic patients, who suffer from disturbed social interactions, present defects in emotional, prosodic, and mentalizing processes (Brüne 2005).

Notes

This work was supported by a grant of the Basse-Normandie Regional Council. We would like to thank Guy Perchey for his help during data acquisitions, Marc Joliot, Frank Lamberton, and Nicolas Delcroix for the analysis of the data, and Gaël Jobard for his precious comments on the manuscript. *Conflict of Interest:* None declared.

Address correspondence to Nathalie Tzourio-Mazoyer, UMR6194 GIP Cyceron, BP 5229, 14074 Caen Cedex, France. Email: tzourio@cyceron.fr.

Appendix A: Examples of the Corpus' Sentences

Emotional Sentences

Anger: J'ai encore retrouvé ma voiture neuve toute rayée, c'est inadmissible/I have found my car with scratch, this is unacceptable.

Happiness: J'ai eu tous mes partiels en juin/I succeeded my entire exam in June.

Sadness: J'ai finalement compris que je ne la reverrais plus/I finally realized that I would never see her again.

Attitudinal Sentences

Doubt: Tu crois vraiment que c'est lui qui a fait ça/Are you sure that he did it?

Irony: J'ai adoré la douceur de ce livre d'horreur/I appreciated the sweetness of this horror book.

Obviousness: Avant de prendre l'avion, j'ai acheté les billets/Before taking a plane, I bought a ticket.

Neutral Sentences

La carafe est remplie de jus d'orange/The carafe is filled with orange juice.

Le cheval court dans la prairie/The horse runs in the meadow.

References

- Ackermann H, Riecker A. 2004. The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain Lang* 89:320-328.
- Adams RB, Janata P. 2002. A comparison of neural circuits underlying auditory and visual object categorization. *Neuroimage* 16:361-377.
- Adolphs R. 2002. Neural systems for recognizing emotion. *Curr Opin Neurobiol* 12:169-177.
- Ashburner J, Friston KJ. 1999. Nonlinear spatial normalization using basis functions. *J Acoust Soc Am* 106:449-457.
- Beauregard M, Chertkow H, Bub D, Murtha S, Dixon R, Evans A. 1997. The neural substrates for concrete, abstract, and emotional word lexica: a positron emission tomography. *J Cogn Neurosci* 9:441-461.
- Belin P, Fecteau S, Bedard C. 2004. Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci* 8:129-135.
- Belin P, Zatorre RJ. 2003. Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 16:2105-2109.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. *Nature* 403:309-312.
- Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y. 1998. Lateralization of speech and auditory temporal processing. *J Cogn Neurosci* 10:536-540.
- Benson RR, Whalen DH, Richardson M, Swainson B, Clark VP, Lai S, Liberman AM. 2001. Parametrically dissociating speech and non-speech perception in the brain using fMRI. *Brain Lang* 78:364-396.
- Blair RJ, Morris JS, Frith CD, Perrett DI, Dolan RJ. 1999. Dissociable neural responses to facial expressions of sadness and anger. *Brain* 122(Pt 5):883-893.
- Blonder LX, Bowers D, Heilman KM. 1991. The role of the right hemisphere in emotional communication. *Brain* 114(Pt 3):1115-1127.
- Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM. 2002. Modality independence of word comprehension. *Hum Brain Mapp* 16:251-261.
- Botvinick MM, Cohen JD, Carter CS. 2004. Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn Sci* 8:539-546.
- Bowers D, Coslett HB, Bauer RM, Speedie LJ, Heilman KM. 1987. Comprehension of emotional prosody following unilateral hemispheric lesions: processing defect versus distraction defect. *Neuropsychologia* 25:317-328.
- Brüne M. 2005. "Theory of mind" in schizophrenia: a review of the literature. *Schizophr Bull* 31:21-42.

- Brunet E, Sarfati Y, Hardy-Bayle MC, Decety J. 2000. A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage* 11:157-166.
- Bush G, Luu P, Posner MI. 2000. Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci* 4:215-222.
- Castelli F, Happe 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.
- Collins DL, Neelin P, Peters TM, Evans AC. 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assisted Tomogr* 18:192-205.
- Content A, Mousty P, Radeau M. 1990. BRULEX: une base de données lexicales informatisée pour le français écrit et parlé. *Année Psychol* 90:551-566.
- Craik FIM, Moroz TM, Moscovitch M, Stuss DT, Winocur G, Tulving E, Kapur S. 1999. In search of the self: a positron emission tomography study. *Psychol Sci* 10:26-34.
- den Ouden HE, Frith U, Frith C, Blakemore SJ. 2005. Thinking about intentions. *Neuroimage* 28:787-796.
- Embick D, Marantz A, Miyashita Y, O'Neil W, Sakai KL. 2000. A syntactic specialization for Broca's area. *Proc Natl Acad Sci USA* 97:6150-6154.
- Fecteau S, Armony JL, Joannette Y, Belin P. 2004a. Priming of non-speech vocalizations in male adults: the influence of the speaker's gender. *Brain Cogn* 55:300-302.
- Fecteau S, Armony JL, Joannette Y, Belin P. 2004b. Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* 23:840-848.
- Ferstl EC, Von Cramon DY. 2002. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *Neuroimage* 17:1599-1612.
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD. 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* 57:109-128.
- Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RSJ. 1995. Spatial registration and normalization of images. *Hum Brain Mapp* 2:165-189.
- Gallagher HL, Frith CD. 2003. Functional imaging of 'theory of mind'. *Trends Cogn Sci* 7:77-83.
- Gallagher HL, Happe F, Brunswick N, Fletcher PC, Frith U, Frith CD. 2000. Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38:11-21.
- George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P, Post RM. 1996. Understanding emotional prosody activates right hemisphere regions. *Arch Neurol* 53:665-670.
- Gervais H, Belin P, Boddaert N, Leboyer M, Coez A, Sfaello I, Barthelemy C, Brunelle F, Samson Y, Zilbovicius M. 2004. Abnormal cortical voice processing in autism. *Nat Neurosci* 7:801-802.
- Goel V, Gold B, Kapur S, Houle S. 1998. Neuroanatomical correlates of human reasoning. *J Cogn Neurosci* 10:293-302.
- Goel V, Grafman J, Sadato N, Hallett M. 1995. Modeling other minds. *Neuroreport* 6:1741-1746.
- Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat Neurosci* 8:145-146.
- Griffiths TD, Buchel C, Frackowiak RS, Patterson RD. 1998. Analysis of temporal structure in sound by the human brain. *Nat Neurosci* 1:422-427.
- Harris LT, Todorov A, Fiske ST. 2005. Attributions on the brain: neuroimaging dispositional inferences, beyond theory of mind. *Neuroimage* 28:763-769.
- Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4:131-138.
- Homae F, Hashimoto R, Nakajima K, Miyashita Y, Sakai KL. 2002. From perception to sentence comprehension: the convergence of auditory and visual information of language in the left inferior frontal cortex. *Neuroimage* 16:883-900.
- Hynes CA, Baird AA, Grafton ST. 2006. Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia* 44:374-383.
- Iacoboni M, Lieberman MD, Knowlton BJ, Molnar-Szakacs I, Moritz M, Throop CJ, Fiske AP. 2004. Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage* 21:1167-1173.
- Jobard G, Crivello F, Tzourio-Mazoyer N. 2003. Evaluation of the dual route theory of reading: a meta-analysis of 35 neuroimaging studies. *Neuroimage* 20:693-712.
- Johnson SC, Schmitz TW, Kawahara-Baccus TN, Rowley HA, Alexander AL, Lee J, Davidson RJ. 2005. The cerebral response during subjective choice with and without self-reference. *J Cogn Neurosci* 17:1897-1906.
- Kircher TT, Brammer M, Tous AN, Williams SC, McGuire PK. 2001. Engagement of right temporal cortex during processing of linguistic context. *Neuropsychologia* 39:798-809.
- Kotz SA, Meyer M, Alter K, Besson M, Von Cramon DY, Friederici AD. 2003. On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang* 86:366-376.
- Kriegstein K, Eger E, Kleinschmidt A, Giraud A. 2003. Modulation of neural responses to speech by directing attention to voices or verbal content. *Cogn Brain Res* 17:48-55.
- Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. 2005. Interaction of face and voice areas during speaker recognition. *J Cogn Neurosci* 17:367-376.
- Kuperberg GR, McGuire PK, Bullmore ET, Brammer MJ, Rabe-Hesketh S, Wright IC, Lythgoe DJ, Williams SCR, David AS. 2000. Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: an fMRI study. *J Cogn Neurosci* 12:321-341.
- Lakshminarayanan K, Ben Shalom D, van Wassenhove V, Orbelo D, Houde J, Poeppel D. 2003. The effect of spectral manipulations on the identification of affective and linguistic prosody. *Brain Lang* 84:250-263.
- Lane RD, Reiman EM, Ahern GL, Schwartz GE, Davidson RJ. 1997. Neuroanatomical correlates of happiness, sadness, and disgust. *Am J Psychiatry* 154:926-933.
- Lawrence EJ, Shaw P, Giampietro VP, Surguladze S, Brammer MJ, David AS. 2006. The role of 'shared representations' in social perception and empathy: an fMRI study. *Neuroimage* 29:1173-1184.
- Lieberman AM, Whalen DH. 2000. On the relation of speech to language. *Trends Cogn Sci* 4:187-196.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150-157.
- Luke KK, Liu HL, Wai YY, Wan YL, Tan LH. 2002. Functional anatomy of syntactic and semantic processing in language comprehension. *Hum Brain Mapp* 16:133-145.
- Mesulam MM. 2000. Behavioral neuroanatomy. In: Mesulam MM, editor. *Principles of behavioral and cognitive neurology*. Oxford: Oxford University Press. p. 1-95.
- Meyer M, Alter K, Friederici AD, Lohmann G, Von Cramon DY. 2002. fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17:73-88.
- Mitchell JP, Banaji MR, Macrae CN. 2005. The link between social cognition and self-referential thought in the medial prefrontal cortex. *J Cogn Neurosci* 17:1306-1315.
- Mitchell JP, Macrae CN, Banaji MR. 2005a. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *J Neurosci* 24:4912-4917.
- Mitchell JP, Macrae CN, Banaji MR. 2005b. Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage* 26:251-257.
- Mitchell RL, Elliott R, Barry M, Cruttenden A, Woodruff PW. 2003. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia* 41:1410-1421.
- Morel M, Lacheret-Dujour A. 2001. "Kali", synthèse vocale à partir du texte: de la conception à la mise en oeuvre. *Traitement Automatique Langues* 42:1-29.
- Northoff G, Bermpohl F. 2004. Cortical midline structures and the self. *Trends Cogn Sci* 8:102-107.

- Ochsner KN, Beer JS, Robertson ER, Cooper JC, Gabrieli JD, Kihlstrom JP, D'Esposito M. 2005. The neural correlates of direct and reflected self-knowledge. *Neuroimage* 28:797-814.
- Ochsner KN, Bunge SA, Gross JJ, Gabrieli JD. 2002. Rethinking feelings: an fMRI study of the cognitive regulation of emotion. *J Cogn Neurosci* 14:1215-1229.
- Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97-113.
- Phan KL, Wager T, Taylor SF, Liberzon I. 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16:331-348.
- Pierce K, Muller RA, Ambrose J, Allen G, Courchesne E. 2001. Face processing occurs outside the fusiform 'face area' in autism: evidence from functional MRI. *Brain* 124:2059-2073.
- Pihan H, Altenmuller E, Ackermann H. 1997. The cortical processing of perceived emotion: a DC-potential study on affective speech prosody. *Neuroreport* 8:623-627.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD. 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10:15-35.
- Premack D, Woodruff G. 1978. Does the chimpanzee have a theory of mind? *Behav Brain Sci* 4:515-526.
- Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RS, Friston KJ. 1996. Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain* 119(Pt 3): 919-931.
- Reiman EM, Lane RD, Ahern GL, Schwartz GE, Davidson RJ, Friston KJ, Yun LS, Chen K. 1997. Neuroanatomical correlates of externally and internally generated human emotion. *Am J Psychiatry* 154:918-925.
- Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. 2004. The role of the medial frontal cortex in cognitive control. *Science* 306:443-447.
- Ross ED. 1981. The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Arch Neurol* 38:561-569.
- Ross ED, Thompson RD, Yenkosky J. 1997. Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain Lang* 56:27-54.
- Ruby P, Decety J. 2003. What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur J Neurosci* 17:2475-2480.
- Ruby P, Decety J. 2004. How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J Cogn Neurosci* 16:988-999.
- Rushworth MF, Walton ME, Kennerley SW, Bannerman DM. 2004. Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci* 8:410-417.
- 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.
- Schmitz TW, Johnson SC. 2005. Self-appraisal decisions evoke dissociated dorsal-ventral aMPFC networks. *Neuroimage*. Forthcoming.
- Schultz RT. 2005. Developmental deficits in social perception in autism: the role of the amygdala and fusiform area. *Int J Dev Neurosci* 23:125-141.
- Scott SK, Young AW, Calder AJ, Hellawell DJ, Aggleton JP, Johnson M. 1997. Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385:254-257.
- Sugiura M, Gotoh R, Okada K, Yamaguchi K, Itoh M, Fukuda H, Kawashima R. 2004. Target dependency of brain mechanism involved in dispositional inference: a PET study. *Neuroimage* 21: 1377-1386.
- Tzourio N, Massiou FE, Crivello F, Joliot M, Renault B, Mazoyer B. 1997. Functional anatomy of human auditory attention studied with PET. *Neuroimage* 5:63-77.
- Verard L, Allain P, Traverso JM, Baron JC, Bloyet D. 1997. Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis. *IEEE Trans Med Imaging* 16:610-616.
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N. 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*. Forthcoming.
- Vogeley K, Bussfeld P, Newen A, Herrmann S, Happe F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K. 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage* 14:170-181.
- Vollm BA, Taylor AN, Richardson P, Corcoran R, Stirling J, McKie S, Deakin JF, Elliott R. 2006. Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a non-verbal task. *Neuroimage* 29:90-98.
- Wagner AD, Pare-Blagoev EJ, Clark J, Poldrack RA. 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31:329-338.
- Walter H, Adenzato M, Ciaramidaro A, Enrici I, Pia L, Bara BG. 2004. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J Cogn Neurosci* 16:1854-1863.
- Wildgruber D, Hertrich I, Riecker A, Erb M, Anders S, Grodd W, Ackermann H. 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cereb Cortex* 14:1384-1389.
- Wildgruber D, Pihan H, Ackermann H, Erb M, Grodd W. 2002. Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage* 15:856-869.
- Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H. 2005. Identification of emotional intonation evaluated by fMRI. *Neuroimage* 24:1233-1241.
- Woods RP, Cherry SR, Mazziotta JC. 1992. Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assisted Tomogr* 16:620-633.