ELSEVIER
NeuroImage
www.elsevier.com/locate/ynimg
NeuroImage xx (2004) xxx–xxx

Cerebral mechanisms of prosodic integration: evidence from connected speech

Isabelle Hesling,[a,*] Sylvain Clément,[b] Martine Bordessoules,[a] and Michèle Allard[a]

[a] *Laboratoire d'Imagerie Moléculaire et Fonctionnelle ERT CNRS, 146 rue Léo Saignat, case 117, 33 076, Université Victor Ségalen, Bordeaux 2, France*
[b] *CNRS UMR 5543, Université Victor Ségalen, Bordeaux 2, France*

Received 20 November 2003; revised 25 June 2004; accepted 1 November 2004

Using functional Magnetic Resonance Imaging (fMRI) and long connected speech stimuli, we addressed the question of neuronal networks involved in prosodic integration by comparing (1) differences in brain activity when hearing connected speech stimuli with high and low degrees of prosodic expression; (2) differences in brain activity in two different diotic listening conditions (normal speech delivery to both ears, i.e., NN; and low-pass-filtered speech delivery to both ears, i.e., FF); and (3) effects of high and low degrees of prosodic information in the NN and FF conditions. Twelve right-handed French men listened passively to the stimuli. Each stimulus induced a specific cerebral network, the flat one weakening activations, which were mainly reduced to the bilateral STG for both listening conditions. High degrees of prosodic information were found to trigger right specific activations in a wider neuronal network involved in speech integration (such as BA44, BA21–22 and BA39–40) than low degrees of prosodic information did. More precisely, the right BA44 was found to be specifically involved in the process of F_0 modulations, which are the main acoustic correlate of prosody. Not only do the results achieved in the present experiment using 30-s-long connected speech stimuli show the involvement of a bilateral neuronal network but they also strongly suggest that high degrees of prosodic information elicit activations in a wider neuronal network involved in speech perception than low degrees of prosodic information do.

Keywords: Connected speech perception; Low-pass filtered speech; Pitch modulations; Hemispheric lateralization; Neuronal network; fMRI

Introduction

The generic term "prosody" has been described as "the organizational structure of speech" (Beckman, 1996) and the prosodic organization of human communication has been reported to be continuous and highly correlated with the semantic, syntactic, morphological and segmental organization of speech (Bolinger, 1972). The emotional relevance of a spoken message is conveyed at once by its semantic content ("what" is said) and by the affective prosody used by the speaker ("how" it is said). From a linguistic point of view, the generic term "prosody" has both a linguistic and an affective content, which leads to the distinction between linguistic prosody–which belongs to the language proper (e.g., in English, different word stresses may distinguish the noun from the verb)—and emotional prosody, which depends on the speaker's emotional state. Emotional prosody involves the expression of emotion through the intonation of spoken language.

Regardless of function, the three main acoustic correlates of the two prosodic phenomena are variations in the fundamental frequency (F_0) or pitch, duration and intensity. Although all types of speech information are communicated simultaneously, the neuroanatomical constituents involved in the processing of the affective quality of the semantic information have been reported to be lateralized in the human brain. A key question is whether the emotional/linguistic prosody dichotomy established in linguistics can be a model for brain–behavior relationships, i.e., for the cerebral processing of prosody. The recent approach of Sidtis and Van Lancker-Sidtis (2003) concludes that prosodic behaviors are coordinated by a complex combination of motor, perceptual and superordinate organizational factors, there ensues that such a complex mechanism can only depend on the recruitment of a complex and largely distributed cerebral network. Indeed, many neuroimaging studies interested in brain–language relationships that have sought to delineate the neural substrate underlying the perception and production of speech prosody, suggest that prosody involves a widespread, bilateral, intertwined and parallel working network. While a broadly distributed network involving the entire perisylvian cortex in the left hemisphere has been reported to play a major role in linguistic processing in most right-handed individuals (Binder et al., 1994; Caplan et al., 1998; Démonet et al., 1992; Friederici et al., 2000; Mazoyer et al., 1993), many neuroimaging studies have supported the notion of a right-hemispheric dominance for processing the affective information conveyed by the voice (for review, see Baum and Pell, 1999; Barrett et al.,

---

[*] Corresponding author. Fax: +33 5 56 79 61 79.
E-mail address: isabelle.hesling@wanadoo.fr (I. Hesling).

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.11.003

1999; Kawashima et al., 1993; Luks et al., 1998; Mayer et al., 2001; Pihan et al., 2000; Stirling et al., 2000).

Closer investigation, either among brain-damaged patients or using neuroimaging techniques on healthy subjects, has yielded four main hypotheses concerning the localization of prosody: (1) the functionalist hypothesis posits a left lateralization of linguistic prosody, in line with the generally held position that language per se is left-lateralized, whereas affective prosody is described as right-lateralized, which fits with generally held positions about emotional processing (Baum, 1998; Van Lancker, 1980); (2) the comprehension and production of prosody are largely subserved by subcortical regions and are not lateralized (Cancelliere and Kertesz, 1990); (3) all aspects of prosody, i.e., the suprasegmental aspects of speech, are processed in the right hemisphere, and the information is then sent through the corpus callosum to the left hemisphere where it is associated with linguistic information, i.e., the segmental aspects of speech (Klouda et al., 1988); (4) individual acoustic cues are reported to be lateralized (Behrens, 1989; Fitch et al., 1997; Gandour et al., 1995; Van Lancker and Sidtis, 1992). Converging evidence strongly supports this cue-dependent hypothesis by suggesting a preferential involvement of the left hemisphere in rapid durational processing (Van Lancker and Sidtis, 1992) as well as a preferential involvement of the right hemisphere in fine spectral processing such as variations of the fundamental frequency or their perceptual correlate, pitch modulations (Cancelliere and Kertesz, 1990; Ivry and Robertson, 1998; Kreiman and Van Lancker, 1988; Pell and Baum, 1997; Poeppel, 2001; Sidtis, 1980, 1984; Sidtis and Feldmann, 1990; Sidtis and Volpe, 1988; Van Lancker and Sidtis, 1992; Zatorre and Belin, 2001; Zatorre et al., 1992, 2002). Nonetheless, this cue-dependent hypothesis does not support the cerebral processing dichotomy of emotional/linguistic prosody since what is called emotional prosody in linguistics is transmitted by many different acoustic cues (such as pitch cues or timing cues), which are processed by different cortical and subcortical areas, thus reflecting the speech redundancy phenomenon. Nevertheless, beyond this linguistic/emotional prosodic debate, there is independent evidence throughout a plethora of studies that the right auditory cortical fields preferentially process pitch variations (Sidtis and Feldmann, 1990; Sidtis and Volpe, 1988; Zatorre, 1988) whereas a bilateral auditory involvement has been reported for pitch processing per se (Griffiths et al., 1998). Moreover, F_0 variations, which are one of the acoustic correlates of pitch variations, are to be considered an important cue for prosody (Blonder et al., 1991; Grosjean and Hirt, 1996; Starkstein et al., 1994; Tompkins and Flowers, 1985; Zatorre, 1988). The emergence of discrepant results may find two convergent explanations: (1) emotional prosody is probably much more complex than the mere linguistic/emotional prosodic dichotomy theory adhering to strict notions of hemispheric lateralization of functions seems to imply, and (2) the heterogeneity of lesions among either right or left brain-damaged patients reveals multiple etiologies of dysprosody (Bryan, 1989; Pell and Baum, 1997; Perkins et al., 1996; Ross et al., 1997; Sidtis and Van Lancker-Sidtis, 2003; Van Lancker and Sidtis, 1992). Consequently, studies using functional imaging techniques on healthy subjects appear to be more reliable for determining anatomical substratum involved in prosodic processing.

In normal subjects, it can be hypothesized that, contrary to nonprosodic information, prosodic information in general, mainly carried by F_0 modulations, is processed in both hemispheres with a preferential involvement of the right auditory cortex. In fact, linguistic prosody and emotional prosody are so embedded that isolating them is all but impossible. In addition, it can be hypothesized that the specific filtering procedure consisting in presenting binaurally low-frequency bands (i.e., modulations of F_0) of prosodic connected speech stimuli induces a much more widely and intertwined network than presentation of flat connected speech stimuli in the same conditions.

In many neuroimaging studies referring to “prosodic speech” vs. “normal speech”, what is called “prosodic speech” is actually a speech signal that has been filtered so as to eliminate lexical, segmental and semantic information (Meyer et al., 2002, 2003; Kotz et al., 2003). Nonetheless, the phrase “prosodic speech” may be misleading in such instances since it is unclear whether the remaining auditory stimuli present high degrees of prosodic information or not. Moreover, since speech exists over time, long connected speech stimuli appear to favor a better integration of pitch modulations, since they present much more F_0 modulations than isolated words or sentences do. That is why we addressed the question of the anatomical substratum involved in prosodic integration when hearing (1) long connected speech and (2) high degrees of prosodic information.

In the present study, we used functional Magnetic Resonance Imaging (fMRI) and two 30-s-long connected speech stimuli differing in degrees of prosodic information (high and low degrees) to assess the neuronal networks involved in prosodic integration. Results were analyzed in order to (1) examine differences produced in brain activity by connected speech stimuli with relatively high degrees of prosodic expression and by connected speech stimuli with low degrees of prosodic expression; (2) compare these differences in brain activity in two diotic listening conditions (normal speech delivery to both ears, i.e., NN; and low-pass filtered speech delivery to both ears, i.e., FF); and (3) effects of high and low degrees of prosodic information in the NN and FF conditions.

Materials and methods

Subjects

Twelve right-handed French men (aged 24–38 years, mean 28.6) participated in the study after giving informed written consent in accordance with the guidelines approved by the Ethics Committee of the Bordeaux Medical University. Subjects had no hearing or neurological disorders and normal structural MRI scans. They had no prior experience of the task and were not familiar with the stimulus material.

Stimuli

Subjects were submitted to two different stimuli each presented in two different experimental listening conditions. Two different 30-s-long connected speech stimuli were chosen, both in French: a very expressive recording and a flat one (the flat speech presentation was purposely different in content from the expressive one in order to prevent subjects from retrieving any prosodic features from the former expressive speech stimulus).

The expressive speech stimulus used a recording of a passage from Edmond Rostand’s drama *Cyrano de Bergerac* and presented pitch modulations ranging from 75 to 300 Hz (Fig. 1A).

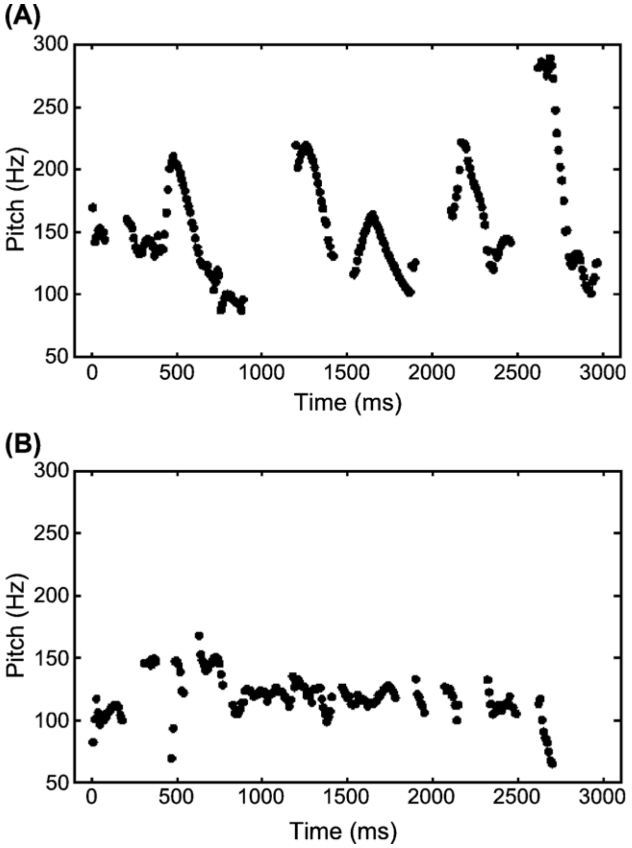

Fig. 1. Pitch modulations of the first 3 s extracted with Wave Surfer software (Sjölander and Beskow, 2000). The upper image (A) illustrates the pitch contour of the expressive speech presentation. The lower image (B) illustrates the pitch contour of the toneless speech presentation.

The flat speech stimulus used a recording of a passage from an economic newspaper, and presented pitch modulations ranging from 70 to 150 Hz (Fig. 1B).

The connected speech stimuli were addressed in four listening conditions and the recording was performed by a trained male speaker in a soundproof room (IAC) at 16bits/41.1 kHz sampling rate. To achieve an acoustic signal exclusively reduced to its prosodic information, the original connected speech was low-pass-filtered using a 10-pole Butterworth filter (using MATLAB DSP Toolbox) providing a 60 dB/oct attenuation, with a 300 Hz cutoff frequency. From a linguistic point of view, the signal derived from this filtering procedure comprised F_0 contour and amplitude envelope, which represent speech melody (distribution and type of pitch accents and boundary markers of prosodic domains). However, acoustic cues largely depend on the lexical and syntactical aspects of the utterances. That is why the two connected speeches not only differed in pitch modulations, but also in rhythm, speech rate and stress, and why their respective acoustic correlates also differed (to a lesser extent) in duration and intensity).

The original and the filtered connected speech productions were presented as follows in four listening conditions for either connected speech stimuli (expressive and flat):

NN: normal speech stimuli (unfiltered speech) presented to both ears,
FN: filtered speech stimuli presented to the left ear and simultaneously unfiltered speech stimuli presented to the right ear,
NF: unfiltered speech stimuli to the left ear and simultaneously filtered speech stimuli to the right ear,
FF: same low-pass-filtered speech stimuli presented to both ears.

The present study restricts the analysis to the diotic presentations, i.e., NN and FF. However, the experimental paradigm and acquisition data are described such as they were performed, i.e., taking into account the four listening conditions.

The FF speech stimuli, which can be described as sounding like a speech listened to from behind a door, was reported to be incomprehensible in the flat speech condition (Table 1).

Procedure

Participants were addressed sixteen 30-s-long stimuli and were asked to listen to the stimuli while remaining motionless and keeping their eyes closed. For each speech presentation, each listening condition was repeated twice as follows: FF-FN-NF-NN-FN-FF-NF-NN. Half of the participants were first presented with the expressive speech stimulus whereas the other half first received the flat one. The speech stimuli were presented binaurally through headphones specifically designed for use in the scanner. To allow for an epoch-related data analysis, all successive presentations of each listening condition lasting 30 s (10 dynamic scans per slice) were separated by an Inter-Trial-Interval of rest lasting 18 s (6 dynamic scans per slice). Then, the hemodynamic response was allowed to return to baseline level. Each speech stimulus was thus composed of nine intertrial intervals of rest and 8 activation phases (i.e., 134 dynamic scans each), the total length of the procedure being 402 s.

Data acquisition

MRI data were collected at 1.5 T using a Gyroscan ACS NT Power track 6000 (Philips Medical System, Best, Netherlands) equipped with echo planar imaging capabilities. Each subject underwent a high-resolution T1-weighted anatomic scan including 30 slices parallel to AC–PC (Anterior Commissure–Posterior Commissure) covering the whole cortex (no gap, thickness 4 mm). Acquisition parameters: TR/TE = 274/25 ms; matrix = 256 × 256; field of view (FOV) = 260 mm. The fMRI data were processed using T2*-weighted gradient echo, echo-planar sequence. Thirty slices (parallel to CA-PC, no gap, thickness 4 mm) were acquired during each TR, acquisition parameters: TR/TE = 3000/60 ms, flip angle = 90°; matrix = 64 × 64; FOV = 260 mm. One hundred thirty-four images per slice were acquired per session.

Table 1
Subjects' performances (shadowing procedure) during NN and FF connected speech presentations outside the scanner

Connected speech presentations	Percent of correct shadowing
Expressive NN connected speech presentation	100
Expressive FF connected speech presentation	60
Flat NN connected speech presentation	100
Flat FF connected speech presentation	15

Data analysis

All data were transferred to a workstation and analyzed using Statistical Parametric Mapping software (Wellcome Department of Cognitive Neurology, London UK, SPM 99) for image realignment and for creation of statistical maps of significant relative regional BOLD (blood oxygenation level dependent) response changes (Friston et al., 1995). Statistical analysis was performed using a boxcar model convoluted by hemodynamic response.

Individual analysis

For both connected speech stimuli, the functional activation of each listening condition was calculated separately, and the interstimulus interval (ISI, i.e., the resting period) after each single trial presentation served as the baseline for data analysis. Voxels were identified as significant only when passing a height threshold of $T = 3$ ($P < 0.001$, uncorrected for multiple comparison) and a minimal 3 voxels was selected per ROI for cluster analysis (192 mm^3). Activation maps were color-coded according to the statistical significance of difference between rest and activation conditions, and then overlaid upon the anatomic T1-weighted images for anatomic reference. The purpose of the individual analysis was (1) to aid anatomical localization of significant activations from the group analysis and (2) to compare the individual activation patterns among subjects with those obtained in the group analysis. Thus, 10 different activated areas were assigned by two experimenters (IH and MA) using Brodmann's classification. An ANOVA was performed considering three factors: listening conditions, hemisphere and region, and the number of activated voxels in each region was considered as the dependent variable. The listening conditions (2 levels) were a between-listening conditions factor and hemisphere (2 levels) and region (10 levels) were within-listening condition group factors. A paired parametric t test was performed whenever activations specifically related to (1) each Region in each Hemisphere (within strategy analysis) and (2) each Region for each listening condition (interstrategy analysis) were significant.

Group analysis

The AC and PC points were identified and functional data sets from each subject were smoothed (Gaussian kernel of 10 mm) and normalized into a standard stereotaxic space (Talairach and Tournoux, 1988), using both the intercommissural line as the reference plane for the transformation and a representative brain from the Montreal Neurological Institute series as template. Areas were thereafter defined both in terms of standard stereotaxic coordinates in the x, y, and z axes and in Brodmann's areas. Data were analyzed using a Random Effects Model for group analysis with a height threshold of $T = 3$ ($P < 0.006$, uncorrected for multiple comparison) and a minimal 20 voxels was selected per ROI for cluster analysis (1280 mm^3).

Then, two-sample t tests were performed between the two original connected speech stimuli (expressive and flat) for each listening condition: expressive NN/ISI vs. flat NN/ISI and expressive FF/ISI vs. flat FF/ISI, the interstimulus interval (ISI) being the baseline. In addition, two-sample t tests were performed between the two listening conditions for the same connected speech stimulus: Expressive NN/ISI vs. FF/ISI and flat NN/ISI vs. FF/ISI. As no assumptions were made about the direction of effects in the condition contrasts, all contrasts were tested and reported for both directions.

Results

Behavioral data

Subjects reported to have understood properly the expressive speech stimulus whatever the listening condition. In contrast, concerning the flat speech stimulus, subjects reported to have encountered problems of intelligibility in the FF condition.

Tests of performances of intelligibility were achieved (through the shadowing procedure) after the scanning session (Table 1) and results showed that subjects were able to repeat both NN connected speech stimuli, whereas their performances decreased in the expressive FF presentation and became quite impossible in the flat FF presentation, which was reported to resemble mere noise. As they could not understand anything even outside the scanner in the flat FF presentation, the scanner noise per se cannot be said to account for their misperception of the flat FF stimulus during the scanning session.

fMRI data

Effects of high degrees of prosodic information

Normal speech presentation (NN). Group analysis revealed that (1) each connected speech stimulus induced a specific pattern of cerebral activation and (2) prosodic utterances produced more activity that flat ones (Figs. 2A,B; Table 2). Whatever the connected speech presentations, the temporal cortex (BA 41–42, 21–22) was bilaterally activated and no interhemispheric differences were observed (Table 2). As for flat speech presentation, both individual and group analyses revealed that it suppressed activations of the articulatory loop, while it increased activations in the right lateral prefrontal gyrus (BA 6).

Two-sample t test analyses of effects of expressive NN presentation vs. flat NN presentation, revealed a specific involvement of the right hemisphere such as the inferior prefrontal cortex (BA 44), the supra temporal region (BA 21) and the inferior parietal gyrus (BA 39–40) (Table 4a). The reverse contrast, i.e., flat NN presentation vs. expressive NN presentation, revealed a right involvement of S2, a left hemispheric involvement of the supra temporal region (BA 22), as well as of the superior parietal gyrus (BA 7) (Table 4a).

Filtered speech presentation (FF)

Group analysis revealed that (1) each connected speech stimulus induced a specific pattern of cerebral activation and (2) prosodic utterances produced more activity than flat ones (Figs. 2C,D). The flat FF presentation mainly reduced activations to the temporal cortex, even if they were weaker than those elicited by the expressive FF presentation, and no interhemispheric differences were observed for either presentation. Conversely, the FF expressive presentation induced bilateral activations in the prefrontal cortex (BA 6), and slight additional activations in the right orbitofrontal cortex (BA 9–10) and bilateral activations in the superior parietal gyrus (BA 7) (Table 3).

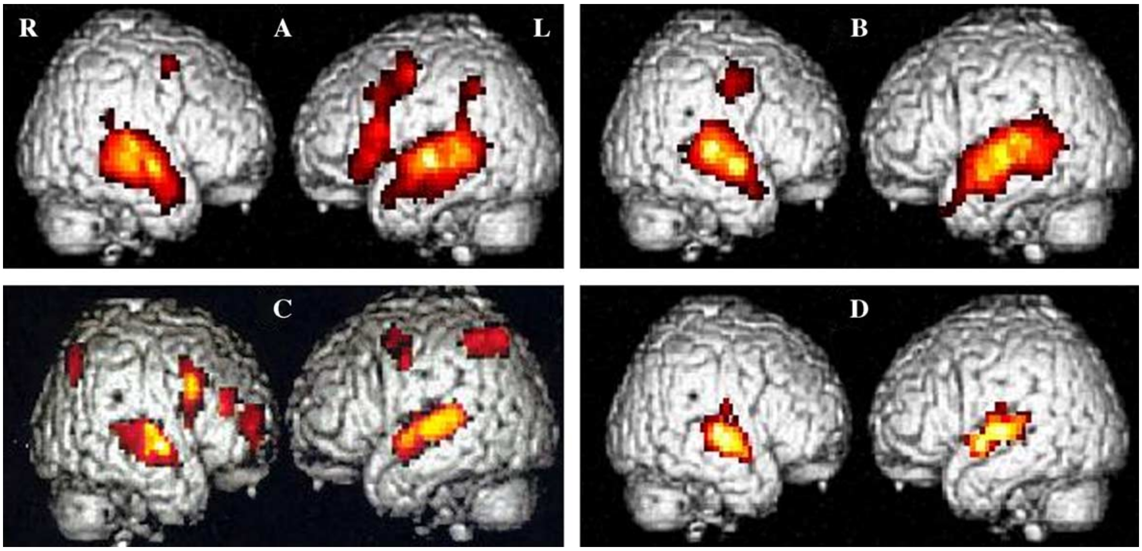

Fig. 2. Patterns of activation elicited by the two following listening conditions: normal speech presentation to both ears (NN); low-pass-filtered speech presentation to both ears (FF), in a prosodic connected speech presentation and a flat one. (A) Expressive speech presentation: normal speech to both ears (NN). Z ranging from 9.54 to 12.54. (B) Expressive speech presentation: low-pass-filtered speech to both ears (FF). Z ranging from 4.06 to 9.96. (C) Flat speech presentation: normal speech to both ears (NN). Z ranging from 5.35 to 13.76. (D) Flat speech presentation: low-pass-filtered speech to both ears (FF). Z ranging from 5.27 to 6.95.

Two-sample t test analyses of the expressive FF presentation vs. the flat FF presentation revealed a right specific involvement of the middle frontal gyrus (BA 9–46) and of the inferior parietal cortex (BA 44) (Table 4b). The reverse contrast, i.e., the flat FF presentation vs. the expressive FF presentation, revealed a left involvement of Heschl's gyrus (BA 41–42), and of the anterior cingulate (BA 24) as well as a right activation of the caudate head (Table 4b).

Table 2
Expressive speech presentation and flat speech presentation (nonfiltered speech to both ears, NN)[a]

Location	BA	K	Z score	Left hemisphere			Right hemisphere		
				x	y	z	x	y	z
Expressive NN presentation									
LPG	6	24	4.86				41	−3	45
LPG	6	243	5.38	−44	−7	42			
IFG	44	244	5.38	−44	−7	37			
HG	41–42	800	5.33				55	−27	5
ITG	38						48	−10	−7
IPG	39–40						55	−44	19
HG	41–42	720	5.38	−54	−21	−5			
STG	21–22			−59	−40	−2			
Flat NN presentation									
LPG	6	88	5.76				50	−6	37
LPG	6	150	5.35	−47	6	41			
STG	22	765	5.55				44	−30	2
ITG	38						50	−10	−7
HG	41–42	833	5.37	−41	−34	7			
STG	22			−41	−24	−1			
ITG	38			−54	−10	−10			
SPG	7	35	4.62	−38	−51	50			

[a] This table and Tables 3–5 list results of loci and significance of each listening condition, with K standing for the number of voxels. Localization is based on stereotaxic coordinates (Talairach and Tournoux, 1988). These coordinates refer to the location of maximal activation indicated by the Z score in a particular anatomical structure. Distances are relative to the intercommissural (AC–PC) line in the horizontal (x), anterior–posterior (y) and vertical (z) directions. The table only lists activation clusters exceeding a minimal size of 20 voxels; LPG, lateral prefrontal gyrus; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; OFG, orbito-frontal gyrus; HG, Heschl's gyrus; STG, superior temporal gyrus; ITG, inferior temporal gyrus; SPG, superior parietal gyrus; IPG, inferior parietal gyrus; aCG, anterior cingulate gyrus.

Table 3
Expressive speech presentation and flat speech presentation (filtered speech to both ears, FF)

Location	BA	*K*	*Z* score	Left hemisphere			Right hemisphere		
				x	*y*	*z*	*x*	*y*	*z*
Expressive FF presentation									
LPG	6	80	4.02				41	1	40
							41	6	28
LPG	6	50	3.11	−41	1	49			
				−44	−10	36			
				−52	−7	33			
MFG	9	56	4.02				50	10	19
MFG	9–46	35	4.28				44	31	20
OFG	10	83	4.28				38	47	8
							32	51	11
HG	41–42	215	4.94				24	−24	5
							41	−21	5
STG	38						50	−17	−5
HG	41–42	330	4.11	−52	−40	4			
				−45	−27	2			
STG	38			−48	−7	−4			
SPG	7	47	3.56				32	−67	39
							32	−75	27
SPG	7	109	4.39	−40	−54	43			
				−30	−67	39			
				−47	−48	43			
Flat FF presentation									
aCG	24	22	3.65	−17	3	49			
HG	41–42	224	6.94				56	−21	15
HG	41–42						50	−13	−4
STG	22						54	−27	−2
HG	41–42	279	6.95	−50	−17	2			
STG	22			−50	−17	15			

Effects of low-pass-filtered speech signal

Expressive connected speech presentation

In the same expressive connected speech presentation, both NN and FF listening conditions induced a specific pattern of activation. The temporal cortex was bilaterally activated in both listening conditions, even if FF activations were weaker. FF suppressed the articulatory loop while it increased activations in the right lateral prefrontal gyrus (BA 6) and elicited new discreet activations in the right orbitofrontal cortex (BA 9–10) as well as bilateral activations in the superior parietal cortex (BA 7).

Two-sample *t* test analyses of expressive NN presentation vs. expressive FF presentation revealed a bilateral involvement of the superior temporal region (right BA 21–22, BA 38 and left BA 21–22) (Table 5a). The reverse contrast revealed a specific right involvement of the middle frontal gyrus (BA 9–46), the inferior prefrontal cortex (BA 44), the orbito-frontal cortex (BA 10) as well as the inferior parietal cortex (BA 39–40) (Table 5a).

The ANOVA with factor hemispheres × listening conditions × regions revealed a main effect of condition ($P < 0.0001$). For each listening condition, no main effect of hemisphere was found. As for the auditory cortex, no significant differences were found between BA 41–42 and BA 21 or BA 22, BA posterior 21 and BA anterior 21, and between BA 21 and BA 22, for each listening condition. Therefore, the comparison between BA posterior 22 and BA anterior 22 yielded different results, which were dependent on the listening condition. The expressive NN condition presented significant differences between the posterior and anterior parts of BA 22 for LH ($P < 0.0101$), whereas expressive FF presented significant differences for RH ($P < 0.0001$), the posterior part of BA 22 being the most activated in both cases.

Flat connected speech presentation

In the same flat connected speech presentation, both NN and FF listening conditions induced a specific pattern of activations. The temporal cortex was bilaterally activated in both listening conditions, though fewer activations were noticed in the supra temporal region for FF. FF elicited a slight new activation in the anterior cingulate gyrus, whereas NN induced bilateral activations in the lateral prefrontal gyrus (BA 6).

Two-sample *t* test analyses revealed a bilateral involvement of the supratemporal region (BA 22–21) (Table 5b). The reverse contrast did not reveal any significant differences between flat FF presentation and expressive FF presentation.

Other statistical analyses performed on individual analysis (ANOVA) revealed no significant differences between the two listening conditions.

Discussion

The major findings of the present study are that (1) bilateral activations of the superior temporal gyrus are observed without any significant left/right differences, whatever the listening conditions and the prosodic nature of the stimulus, (2) prosodic

Table 4
Two-sample t tests

Location	BA	K	Z score	Left hemisphere			Right hemisphere		
				x	y	z	x	y	z
(a) Expressive NN vs. flat NN presentations									
IFG	44	31	3.21				41	20	13
IPG	39–40	29	3.67				44	−40	35
STG	21 post	23	4.49				44	−34	−6
STG	21 ant	33	4.42				47	−3	−22
IPG	39–40	26	3.67	−54	−51	34			
Flat NN vs. expressive NN presentations									
S2		54	3.78				52	−17	36
STG	22 post	50	5.63	−44	−75	−10			
SPG	7	209	4.91	−32	−54	52			
(b) Expressive FF vs. flat FF presentations									
MFG	9–46	89	4.24				47	14	34
IFG	44	20	3.06				38	31	17
Flat FF vs. expressive NN presentations									
HG	41–42	88	4.05	−13	3	12			
aCG	24	35	3.62	−13	24	22			
Caudate		47	3.68				47	−21	15
Caudate		95	3.99				19	−64	−4

(a) Expressive speech presentation, normal speech stimulus to both ears (NN)/interstimulus interval (ISI) vs. flat speech stimulus, normal speech stimulus to both ears (NN)/interstimulus interval (ISI) ($P < 0.001$). (b) Expressive speech presentation, filtered speech stimulus to both ears (FF)/interstimulus interval (ISI) vs. flat speech stimulus, filtered speech stimulus to both ears (FF)/interstimulus interval (ISI) ($P < 0.001$).

connected speech stimuli produce more activity than the flat ones, especially in right areas such as the inferior prefrontal cortex, the superior temporal gyrus, and the inferior parietal gyrus. These results show that even if speech perception involves a very widespread, bilateral and parallel working network in which many different brain areas are closely intertwined, the functional cerebral network largely depends on the prosodic content of the stimuli.

Whatever the listening conditions (NN or FF) and the nature of connected speech stimuli (expressive or flat) that were used, auditory areas subserving speech comprehension were found to be bilaterally activated, which is in accordance with neuroimaging studies on the processing of spoken words (Binder et al., 2000, Mazoyer et al., 1993) and sentences (Dehaene et al., 1997; Kuperberg et al., 2000; Meyer et al., 2002). In addition, neither the order of presentation of the two connected speech stimuli (i.e., whether the expressive speech was presented in first or second place), nor the orders of presentation of the two listening conditions induced any significant differences in the recruited neuronal patterns. Nonetheless, whatever the nature of the auditory stimuli, the NN condition induced more activation in bilateral secondary auditory areas (BA 22–21) than the FF condition did, as revealed by two-sample t test analyses (Table 5). Such an observation may be accounted for by hypothesizing that as the FF condition reduces the recorded speech to a less comprehensible message, the primary auditory cortex is the most activated part of the STG, whereas the secondary auditory cortex, involved in speech comprehension, is less activated. This finding is consistent with a handful of fMRI studies reporting a decrease in activation in auditory cortices when subjects are required to listen to a purely prosodic speech (Meyer et al., 2002) or when they listen to incomprehensible speech (Poldrack et al., 2001). Our results are also in accordance with previous data showing that auditory sentence comprehension is associated with involvement of both left and right STG (Kuperberg et al., 2000; Meyer et al., 2002; Muller et al., 1997; Schlosser et al., 1998). However, in the present experiment, both group analysis and individual analysis revealed that neither expressive nor flat NN presentations induced a left STG dominance, contrary to the findings of previous studies focusing on general syntactic operations at sentence level (Friederici et al., 2000; Humphries et al., 2001; Meyer et al., 2000; Scott et al., 2000). Methodological differences such as the duration of the stimuli, the nature of the stimuli (high and low degrees of prosodic information) as well as the nature of the task may explain this discrepancy in results. More precisely, it has been reported that when acoustic properties are controlled, a left lateralized response to speech can be observed (Scott et al., 2000). But the present experiment did not require any acoustic controls. Nonetheless, a significant difference between expressive NN presentation vs. flat NN presentation could be noticed in the right STR (BA 21) (Table 4a), which leads us to hypothesize that a high degree of prosodic information induces a more important involvement of the right auditory cortex than a low degree of prosodic information does. In fact, as the expressive speech stimulus used in our experiment presented a much wider range of F_0 variations compared with the flat stimulus (Fig. 1), the right STR involvement in this case can be explained by the specific processing of these F_0 variations (or their perceptual correlate pitch), which have been reported to be preferentially processed in the right auditory cortex (Cancelliere and Kertesz, 1990; Ivry and Robertson, 1998; Kreiman and Van Lancker, 1988; Meyer et al., 2002; Pell and Baum, 1997; Poeppel, 2001; Sidtis, 1980, 1984; Sidtis and Feldmann, 1990; Sidtis and Volpe, 1988; Tzourio et al., 1997; Van Lancker and

Table 5
Two-sample t tests

Location	BA	K	Z score	Left hemisphere			Right hemisphere		
				x	y	z	x	y	z
(a) Expressive NN vs. FF presentation									
STG	22–21	511	6.93				50	−20	7
ITG	38	40	3.62				29	−37	15
STG	22–21	640	5.83	−41	−34	4			
Caudate		36	3.74				10	6	4
Expressive FF vs. NN presentation									
MFG	9–46	28	3.10				41	28	26
IFG	44	64	4.24				41	6	37
OFG	10	34	3.56				30	55	6
IPG	39–40	44	3.59				29	−67	33
(b) Flat NN vs. FF presentation									
STG	22–21	538	6.61				34	−30	4
STG	21 post	24	3.82				50	−64	−4
STG	22–21	513	6.10	−47	−34	4			
SPG	7	21	5.66	−30	−58	52			
Flat FF vs. NN presentation									
–	–	–	–	–	–	–	–	–	–

(a) Normal speech stimulus to both ears (NN)/interstimulus interval (ISI) vs. filtered speech to both ears (FF)/interstimulus interval (ISI), in expressive speech presentation ($P < 0.001$). (b) Normal speech stimulus to both ears (NN)/interstimulus Interval (ISI) vs. filtered speech to both ears (FF)/interstimulus interval (ISI), in flat speech presentation ($P < 0.001$).

Sidtis, 1992; Zatorre and Belin, 2001; Zatorre and Sansom, 1991; Zatorre et al., 1992, 2002).

The flat FF condition, which carries only the F_0 modulations, revealed activations mainly reduced to both the left and right primary auditory cortices (Fig. 2, Table 3) since the message, deprived of any prosodic modulations, is then reduced to mere noise, which cannot be decoded by subjects either semantically or prosodically. In addition, the flat FF speech presentation revealed more activation in the left Heschl's gyrus than the expressive FF speech presentation did (Table 4b), which is also consistent with the fact subjects only heard mere noise. A pilot study conducted before the fMRI experiment revealed that when the expressive connected speech stimulus used in the experiment was presented in a flat tone (even if presented before the expressive speech stimulus), subjects reported that by subvocalizing they could not help retrieving the specific prosodic aspects of such a well-known declamation. This assessment led us to propose two speech stimuli differing in their semantic content to make sure that prosodic vs. aprosodic discrimination would be effective. Moreover, as subjects could not understand the flat FF stimulus outside the scanner (Table 1), they were not expected to understand it better during the scanning session. Nonetheless, the flat FF presentation did not recruit any other areas than the areas usually involved in a listening process.

There is ample evidence that speech perception requires such a highly distributed network that many brain regions beyond the traditional language areas of Wernicke and Broca are involved, especially the right prefrontal cortex (BA 6–44). The prefrontal activations achieved in our results vary according to the listening conditions and to the nature of the connected speech stimuli (expressive vs. flat). The expressive NN condition yielded a left prefrontal activation (BA 6–44), which was not achieved in the flat NN (Fig. 2, Table 4a). Nonetheless, the contrast between the two speech stimuli in the NN listening condition did not reveal any activation in this specific area. It can thus be hypothesized that subliminal activations of this specific left prefrontal region occurred in the flat NN presentation. This hypothesis is in accordance with previous work suggesting a functional connection between the auditory cortex and the prefrontal associative cortex involved both in the retrieval and rehearsal of auditory information and in auditory working memory (Buchanan et al., 2000; Démonet et al., 1992; Zatorre et al., 1992, 1996). Broca's area (left BA 44) has also been reported to be activated during speech perception (Caplan et al., 2000; Mazoyer et al., 1993; Paulesu et al., 1993; Poldrack et al., 1999).

Moreover, the right frontal operculum activation (BA 44), achieved both in the expressive NN presentation vs. the flat NN presentation and in the expressive FF presentation vs. the flat FF presentation (Table 4), i.e., in an expressive context compared with a flat one, may be related to durational processes, as revealed by many studies underlining the involvement of this area (as well as of its left counterpart, namely Broca's area) in temporal analyses (Fiez et al., 1995; Platel et al., 1997; Schubotz et al., 2000; Tallal et al., 1993). More specifically, rhythm analysis tasks have revealed either a bilateral prefrontal activation (BA 6–44) (Schubotz et al., 2000) or a right lateralized one (Riecker et al., 2000; Zatorre et al., 1992, 1994). In addition, a right hemispheric dominance for the recognition of the so-called "emotional prosody" has been widely assumed by some neuroimaging studies (Breitenstein et al., 1998; George et al., 1996; Imaizumi et al., 1997; Ross et al., 1997).

Left activations of both the angular gyrus (BA 39) and the supramarginal gyrus (BA 40) only achieved in two-sample t test analyses of effects of expressive NN presentation vs. flat NN presentation (Table 4a) are in accordance with previous neuroimaging studies on speech perception. More precisely, the left inferior parietal lobule has been widely described as playing a role in speech comprehension processing and more particularly as being involved both (1) in the processing of auditory spatial information, the "where" pathway (Rauschecker and Tian, 2000; Weeks et al., 1999; Zatorre et al., 1992), and (2) in the processing of passive phonological store (Paulesu et al., 1993; Shallice and Vellar, 1990; Wildgruber et al., 1999). It can thus be hypothesized that not only do high degrees of prosodic information induce activation of specific brain areas such as the right prefrontal cortex, but they also elicit new activations that are not closely linked to a prosodic process. Conversely, right activations of both the angular gyrus and the supramarginal gyrus appear to be specifically induced by prosodic information, i.e., (1) when the predominant prosodic factor is involved (expressive NN presentation vs. flat NN presentation, (Table 4a) and (2) when F_0 modulations are preferentially involved (expressive FF presentation vs. the expressive NN presentation, (Table 5a). This assumption is strengthened by the observation that this specific area was not recruited when (1) the predominant factor prosody was not involved (flat NN presentation vs. expressive NN presentation (Table 4a), and (2) when F_0 modulations were reduced to a merely flat line (pitch modulations ranging from 70 to 150 Hz) (Table 5b). The right inferior lobule has been described as being activated in the process of phonological information by some neuroimaging studies, even if its involvement is not clearly understood as yet (McDermott et al., 2003; Pillai et al., 2003).

General conclusion

Both the two connected speech stimuli and the two listening conditions used here were found to recruit different neuronal networks. From an information processing perspective, speech perception involves several stages of auditory sensory analysis and pattern extraction as well as interactions between sensory and stored linguistic information. These analytical stages imply a degree of hierarchical functional organization (Binder et al., 2000; Klatt, 1989; Mc Clelland and Elman, 1986). High degrees of prosodic information seem to trigger right specific activations in a wider neuronal network involved in prosodic speech perception (such as the right inferior prefrontal cortex, the right supra temporal gyrus, and the right inferior parietal gyrus) than low degrees of prosodic information do. More precisely, the right inferior prefrontal gyrus (BA 44) seems to be specifically involved in the process of F_0 modulations (or their perceptual correlate pitch), which are the main acoustic correlate of prosody.

Acknowledgments

The authors would like to thank Dr. Laurent Demany for helpful comments on an earlier version of the manuscript; Dr. Didier Cugy and Dr. Bixente Dilharreguy for help with data collection and analysis; and Dr. Ray Cooke as well as Dr. Anne-Marie Carassou for their helpful reading of an earlier version of the manuscript. The research was supported by INSERM (contrat d'interface INSERM/Université Victor Ségalen to IH) and Pôle Aquitaine Santé to MA.

References

Baum, S.R., 1998. The role of fundamental frequency and duration in the perception of linguistic stress by individuals with brain damage. J. Speech Lang. Hear. Res. 41, 31–40.

Baum, S.R., Pell, M.D., 1999. The neural bases of prosody: insights from lesion studies and neuroimaging. Aphasiology 13 (8), 581–608.

Barrett, A.M., Crucian, G.P., Raymer, A.M., Heilman, K.M., 1999. Spared comprehension of emotional prosody in a patient with global aphasia. Neuropsychiatry Neuropsychol. Behav. Neurol. 12, 117–120.

Beckman, M., 1996. The parsing of prosody. Lang. Cogn. Processes 11, 17–68.

Behrens, S., 1989. Characterizing sentence intonation in a right-hemisphere damaged population. Brain Lang. 37, 181–200.

Binder, J.R., Rao, S.M., Hammeke, T.A., 1994. Functional magnetic resonance imaging of human auditory cortex. Ann. Neurol. 35, 662–672.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and non-speech sounds. Cereb. Cortex 10, 512–528.

Blonder, L.X., Bowers, D., Heilman, K.M., 1991. The role of the right hemisphere in emotional communication. Brain 114, 1115–1127.

Bolinger, D., 1972. Intonation: Selected Readings. Penguin Books, Harmondsworth.

Breitenstein, C., Van Lancker, D., Kempler, D., Daum, I., Waters, C.H., 1998. The contribution of working memory to the perception of emotional prosody in Parkinson's disease. Brain Lang. 65, 243–246.

Bryan, K., 1989. Language prosody in the right hemisphere. Aphasiology 3, 285–299.

Buchanan, L., Kiss, I., Burgess, C., 2000. Word and non-word reading in a deep dyslexic: phonological information enhances performance. Brain Cogn. 43, 65–68.

Cancelliere, A., Kertesz, A., 1990. Lesion localization in acquired deficits of emotional expression and comprehension. Brain Cogn. 13, 133–147.

Caplan, D., Alpert, N., Waters, G., 1998. Effects of syntactic structure and propositional number on patterns of regional cerebral flow. J. Cogn. Neurosci. 10, 541–552.

Caplan, D., Alpert, N., Waters, G., Olivieri, A., 2000. Activation of Broca's area by syntactic processing under conditions of concurrent articulation. Hum. Brain Mapp. 9, 65–71.

Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., van de Moortele, P., Lehricy, S., Le Bihan, D., 1997. Anatomical variability in the cortical representation of first and second language. NeuroReport 8, 3809–3815.

Démonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., Frackowiak, R.S.J., 1992. The anatomy of phonological and semantic processing in normal subjects. Brain 115, 1753–1768.

Fiez, J.A., Tallal, P., Raichle, M.E., Katz, W.F., Miezin, F.M., Petersen, S.E., 1995. PET studies of auditory and phonological processing: effects of stimulus type and task condition. J. Cogn. Neurosci. 7, 357–375.

Fitch, R.H., Miller, S., Tallal, P., 1997. Neurobiology of Speech Perception. Annu. Rev. Neurosci. 20, 331–353.

Friederici, A.D., Meyer, M., von Cramon, D.Y., 2000. Auditory language comprehension: an event related fMRI study on the processing of syntactic and lexical information. Brain Lang. 75, 465–477.

Friston, K.J., Ashburner, J., Poline, J.B., Frith, C.D., Heather, J.D., Frackowiak, R.S.J., 1995. Spatial registration and normalization of images. Hum. Brain Mapp. 2, 165–189.

Gandour, J., Larsen, J., Dechongkit, S., Ponglorpisit, S., Khunadorn, F., 1995. Speech prosody in affective contexts in Thai patients with Right-Hemisphere Lesions. Brain Lang. 51 (3), 422–443.

George, M.S., Parekh, P.I., Rosinsky, N., 1996. Understanding emotional prosody activates right hemisphere regions. Arch. Neurol. 53, 665–670.

Griffiths, T.D., Büchel, C., Frackowiak, R.S.J., Patterson, R.D., 1998. Analysis of temporal structure in sound by the human brain. Nat. Neurosci. 5 (1), 422–427.

Grosjean, F., Hirt, C., 1996. Using prosody to predict the end of sentences in English and French: normal and brain-damaged subjects. Lang. Cogn. Processes 11 (1/2), 107–134.

Humphries, C., Buchsbaum, C., Hickok, G., 2001. Role of the anterior temporal cortex in auditory sentence comprehension: an fMRI study. NeuroReport 12, 1749–1752.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., Itoh, K., Kato, T., Nakamura, A., Hatano, K., Kojima, S., Nakamura, K., 1997. Vocal identification of speaker and emotion activates different brain regions. NeuroReport 8, 2809–2812.

Ivry, R., Robertson, L.C., 1998. The Two Sides of Perception. Series in Cognitive Neurosciences. MIT Press, Cambridge.

Kawashima, R., Itoh, M., Hatazawa, J., Miyazawa, H., Yamada, K., Matsuzawa, T., Fukada, H., 1993. Changes of cerebral blood flow during listening to unfamiliar spoken language. Neurosci. Lett. 161, 69–72.

Klatt, D.H., 1989. Review of text-to-speech conversion for English. J. Acoust. Soc. Am. 87, 820–857.

Klouda, G.V., Robin, D.A., Graff-Radford, N.R., Cooper, W., 1988. The role of callosal connections in speech prosody. Brain Lang. 35, 154–171.

Kotz, S.A., Meyer, M., Alter, K., Besson, M., von Cramon, D.Y., Friederici, A.D., 2003. On the lateralization of emotional prosody: an event-related functional MR investigation. Brain Lang. 86, 366–376.

Kreiman, J., Van Lancker, D., 1988. Hemispheric specialization for voice recognition: Evidence from dichotic listening. Brain Lang. 34, 246–252.

Kuperberg, G.R., Mc Guire, P.K., Bullmore, E.T., Brammer, M.J., Rabe-Hesketh, S., Wright, I.C., Lythgoe, D.J., Williams, S.C.R., David, A.S., 2000. Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: an fMRI study. J. Cogn. Neurosci. 12, 321–341.

Luks, T.L., Nusbaum, H.C., Levy, J., 1998. Hemispheric involvement in the perception of syntactic prosody in dynamically dependent on task demands. Brain Lang. 65, 313–332.

Mayer, J., Wildgruber, D., Ackermann, H., Dogil, G., Riecker, A., Grood, W., 2001. Lateralized fMRI activation at the level of the motor cortex during linguistic and affective prosody production tasks. Proceedings of the 7th International Conference on Functional Mapping of the Human Brain, Brighton, June 10–14 2001, NeuroImage, vol. 13, p. S568.

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.

Mc Clelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. Cogn. Psychol. 18, 1–86.

McDermott, K.B., Petersen, S.E., Watson, J.M., Ojemann, J.G., 2003. A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. Neuropsychologia 41 (3), 293–303.

Meyer, M., Friederici, A.D., von Cramon, D.Y., 2000. Neurocognition of auditory sentence comprehension: event-related fMRI reveals sensitivity violations and task demands. Brain Res. Cogn. Brain Res. 9, 19–33.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002. fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum. Brain Mapp. 17 (2), 73–88.

Meyer, M., Alter, K., Friederici, A.D., 2003. Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. J. Neurolinguist. 16, 277–300.

Muller, R.A., Rothermel, R.D., Behen, M.E., Muzik, O., Mangner, T.J., Chugani, H.T., 1997. Receptive and expressive language activations for sentences: a PET study. NeuroReport 8, 3767–3770.

Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal components of working memory. Nature 362, 342–345.

Pell, M.D., Baum, S.R., 1997. The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. Brain Lang. 57, 80–89.

Perkins, J.M., Baran, J.A., Gandour, J., 1996. Hemispheric specialization in processing intonation contours. Aphasiology 10, 343–362.

Pihan, H., Altenmuller, E., Hertrich, I., Ackermann, H., 2000. Cortical activation patterns of affective speech processing depend on concurrent demands on the sub vocal rehearsal system: a DC potential study. Brain 123, 2338–2349.

Pillai, J.J., Araque, J.M., Allison, J.D., Sethuraman, S., Loring, D.W., Thiruvaiyaru, D., Ison, C.B., Balan, A., Lavin, T., 2003. Functional MRI study of semantic and phonological language processing in bilingual subjects: preliminary findings. NeuroImage 19, 565–576.

Platel, H., Price, C., Baron, J.C., Wise, R., Lambert, J., Frackowiak, R.S.J., 1997. The structural components of music perception: a functional anatomical study. Brain 120, 229–243.

Poeppel, D., 2001. Pure word deafness and the bilateral processing of the speech code. Cogn. Sci. 25, 679–693.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrieli, J.D.E., 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. NeuroImage 10, 15–35.

Poldrack, R.A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., Gabrieli, J.D.E., 2001. Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. J. Cogn. Neurosci. 13, 687–697.

Rauschecker, J.P., Tian, B., 2000. Mechanisms and stream for processing “what” and “where” in auditory cortex. PNAS 97, 11800–11806.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., Grodd, W., 2000. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. NeuroReport 11, 1997–2000.

Ross, E.D., Thompson, R.D., Yenkosky, J., 1997. Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. Brain Lang. 56, 27–54.

Schlosser, M.J., Aoyagi, N., Fulbright, R.K., Gore, J.C., Mc Carthy, G., 1998. Functional MRI studies of auditory comprehension. Hum. Brain Mapp. 6, 1–13.

Schubotz, R.I., Friederici, A.D., Von Cramon, D.Y., 2000. Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. NeuroImage 11, 1–12.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.

Shallice, T., Vellar, G., 1990. The impairment of auditory verbal short-term storage. In: Vellar, T., Shallice, T. (Eds.), Neuropsychological Impairments of Short-Term Memory. Cambridge University Press, Cambridge, UK.

Sidtis, J.J., 1980. On the nature of the cortical function underlying right hemisphere auditory perception. Neuropsychologia 18, 321–330.

Sidtis, J.J., 1984. Music, pitch perception and the mechanisms of cortical hearing. Handbook of Cognitive Science. Plenum Press.

Sidtis, J.J., Feldmann, E., 1990. Transient ischemic attacks presenting with a loss of pitch perception. Cortex 26 (3), 469–471.

Sidtis, J.J., Van Lancker-Sidtis, D., 2003. A neurobehavioral approach to dysprosody. Semin. Speech Lang. (24), 93–105.

Sidtis, J.J., Volpe, B.T., 1988. Selective loss of complex-pitch or speech discrimination after unilateral lesions. Brain Lang. 34, 235–245.

Sjölander, K., Beskow, J., 2000. WaveSurfer—An Open Source Speech Tool. Proceedings of ICSLP, Beijing, China.

Starkstein, S.E., Federoff, J.P., Price, T.R., Leiguarda, R.C., Robinson, R.G., 1994. Neuropsychological and neuroradiological correlates of emotional prosody comprehension. Neurology 44, 515–522.

Stirling, J., Cavill, J., Wilkinson, A., 2000. Dichotically presented emotionally intoned words produce laterality differences as a function of localisation tasks. Laterality 5, 363–371.

Talairach, J., Tournoux, P., 1988. Co-Planar stereotaxic atlas of the human brain. Thieme, New York.

Tallal, P., Miller, S., Fitch, R., 1993. Neurobiological basis of speech: a case for the pre-eminence of temporal processing. In: Tallal, P., Galaburda, A.M., Llinas, R., von Euler (Eds.), Temporal Information Processing in the Nervous System: Special Reference to Dyslexia and Dysphasia, Ann. N. Y. Acad. Sci., vol. 682, pp. 27–47.

Tompkins, C.A., Flowers, C.R., 1985. Influence of congruent and incongruent contexts on prosodic mood recognition by brain-damaged adults. Clinical Aphasiology Conference, Ashland, Oregon.

Tzourio, N., Massioui, F.E., Crivello, F., Joliot, M., Renault, B., Mazoyer, B.,

1997. Functional anatomy of human auditory attention studied with PET. NeuroImage 5, 63–77.

Van Lancker, D., 1980. Cerebral lateralization of pitch cues in the linguistic signal. Int. J. Hum. Commun. 13, 227–277.

Van Lancker, D., Sidtis, J.J., 1992. The identification of affective-prosodic stimuli by left-and right-hemisphere-damaged subjects: all errors are not created equal. J. Speech Hear. Res. 35, 963–970.

Weeks, R.A., Aziz-Sultan, A., Bushara, K.O., Tian, B., Wessinger, C.M., Dang, N., Rauschecker, J.P., Hallett, M., 1999. A PET study of human auditory spatial processing. Neurosci. Lett. 252, 155–158.

Wildgruber, D., Kischka, U., Ackermann, H., Klose, U., Grodd, W., 1999. Dynamic pattern of brain activation during sequencing of word strings evaluated by fMRI. Brain Res. Cogn. Brain 7, 285–294.

Zatorre, R.J., 1988. Pitch perception of complex tones and human temporal-lobe function. J. Acoust. Soc. Am. 84, 566–572.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11, 946–953.

Zatorre, R.J., Sansom, S., 1991. Role of the right temporal neocortex in retention of pitch in auditory short-term memory. Brain 114, 2403–2417.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14, 1908–1919.

Zatorre, R.J., Halpern, A.R., Perry, D.W., Meyer, E., Evans, A.C., 1996. Hearing in the mind's ear: a PET investigation of musical imagery and perception. J. Cogn. Neurosci. 8, 29–46.

Zatorre, R.J., Belin, P., Penhunen, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46.