NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 24 (2005) 21 – 33

Neural correlates of switching from auditory to speech perception

Ghislaine Dehaene-Lambertz,[a,b,c,*] Christophe Pallier,[b] Willy Serniclaes,[d] Liliane Sprenger-Charolles,[d] Antoinette Jobert,[b] and Stanislas Dehaene[b]

[a]*Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS, ENS and CNRS UMR 8554), IFR 49, France*
[b]*Unité INSERM 562, Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, Orsay, IFR 49, France*
[c]*Service de Neuro-Pédiatrie, CHU Bicêtre, France*
[d]*CNRS UMR 8606 and Université René Descartes, France*

Received 1 December 2003; revised 22 June 2004; accepted 17 September 2004

Many people exposed to sinewave analogues of speech first report hearing them as electronic glissando and, later, when they switch into a 'speech mode', hearing them as syllables. This perceptual switch modifies their discrimination abilities, enhancing perception of differences that cross phonemic boundaries while diminishing perception of differences within phonemic categories. Using high-density evoked potentials and fMRI in a discrimination paradigm, we studied the changes in brain activity that are related to this change in perception. With ERPs, we observed that phonemic coding is faster than acoustic coding: The electrophysiological mismatch response (MMR) occurred earlier for a phonemic change than for an equivalent acoustic change. The MMR topography was also more asymmetric for a phonemic change than for an acoustic change. In fMRI, activations were also significantly asymmetric, favoring the left hemisphere in both perception modes. Furthermore, switching to the speech mode significantly enhanced activation in the posterior parts of the left superior gyrus and sulcus relative to the non-speech mode. When responses to a change of stimulus were studied, a cluster of voxels in the supramarginal gyrus was activated significantly more by a phonemic change than by an acoustic change. These results demonstrate that phoneme perception in adults relies on a specific and highly efficient left-hemispheric network, which can be activated in top-down fashion when processing ambiguous speech/non-speech stimuli.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Speech perception; Auditory; Non-speech stimuli; ERP; MRI

Introduction

Speech perception is the set of processes by which the listener extracts words from the continuous, rapidly changing, acoustic signal of speech. The advent of machines allowing the recording and visualization of the speech signal has made clear that this is not an obvious problem to solve because the acoustic realizations of a given word can vary greatly depending on speech rate, speaker's voice features, context, etc. Despite their apparent variability, words, and the phonemes that constitute them, are most often effortlessly identified. How is such perceptual constancy achieved? Is it thanks to general auditory mechanisms or thanks to specialized speech decoding processes? Because some animals, such as chinchillas (Kuhl and Miller, 1975) or quails (Kluender et al., 1987), can be taught to discriminate speech sounds much as humans do (i.e., categorically), some authors claim that phonemes are processed by general psychoacoustic mechanisms and that the properties of phoneme perception are essentially dependent on the physiological properties of the auditory system (Kluender and Greenberg, 1989; Stevens and Blumstein, 1981). Others have provided empirical arguments (e.g., the phenomenon of duplex perception) favoring a specialized speech processor (Liberman and Mattingly, 1989; Liberman et al., 1967). Moreover, since speech is the main vector of communication for the human species, it is plausible that evolution has selected special mechanisms to process it (Liberman, 1996; Pinker, 1994). This would explain the dispositions toward speech manifested by infants (Bertoncini et al., 1988; Dehaene-Lambertz and Baillet, 1998).

Elucidating the neural bases of general auditory processes and speech decoding processes in the human brain could shed light on this debate. Finding brain areas that show increased activation when acoustic utterances are treated like speech would lend support to the notion of a specialized speech processor (leaving aside the question of innateness, since in adults, it may have emerged because of extensive experience with speech). A number of brain imaging studies have compared the activations elicited by speech versus non-speech stimuli (e.g., Binder et al., 2000; Burton et al., 2000; Celsis et al., 1999; Jacquemot et al., 2003; Vouloumanos et al., 2001). The findings have been somewhat heterogeneous, probably reflecting the different tasks and types of stimuli used. Furthermore, the acoustic properties of the speech and non-speech stimuli differed. As a result, any reported difference in brain activations could be attributed to differences in the acoustic

[*] Corresponding author. Unité INSERM 562, Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, 4 place du général Leclerc, 91401 Orsay cedex, France. Fax +33 1 69 86 78 16.

E-mail address: ghis@lscp.ehess.fr (G. Dehaene-Lambertz).

Available online on ScienceDirect (www.sciencedirect.com.)

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.09.039

features of the stimuli. Thus, researchers have debated whether some activation reported to be specific to speech in the left posterior superior temporal gyrus might be due to the presence of fast spectral transitions (Zatorre and Binder, 2000).

In the present study, our participants had to perform a simple discrimination task on synthetic sinewave analogues of speech. These acoustic stimuli can be perceived either as non-speech or as speech. Stimuli and instructions were similar during the entire experiment, but the mode of perception changed from non-speech to speech in the middle of the experiment. We used high-density event-related potentials and fMRI to separate processes related to the physical characteristics of the stimuli from those related to the mode of perception. With ERPs, our goal was to determine how quickly acoustic and phonemic processes diverge, while fMRI was used to inform about the localization of these processes.

Method

Subjects

Twelve right-handed young French adults (6 females and 6 males, aged from 20 to 30 years) were tested using ERPs and 19 other subjects were tested with fMRI (12 females and 7 males, aged from 18 to 28 years). Four additional subjects were rejected from the ERP experiment because they presented only one mode of perception. Three of them heard syllables immediately and one was unable to hear syllables and heard only whistles during the entire experiment. All subjects gave their written informed consent. None of them had a history of oral or written language impairment, neurological or psychiatric disease, nor any hearing deficit.

Stimuli

The stimuli were four sinewave analogues of a CV syllable, used in a previous behavioral study (Serniclaes et al., 2001). The vowel was /a/ and the consonant was varied along a place of articulation continuum, generated by an amplitude-weighted sum of sinusoids. The onset of the initial frequency transitions (SIN2 and SIN3), which corresponded to those of the second and third formants in natural speech (F2 and F3), varied from 975 to 1800 Hz in three equal steps of 275 Hz for SIN2 and from 1975 to 3400 Hz in three equal steps of 475 Hz for SIN3. The end frequencies of SIN2 and SIN3 transitions were fixed at 1300 and 2500 Hz, respectively. The stimuli are labeled according to their phonemic identity for French speakers (i.e., depending on whether they were predominantly identified as /ba/ or /da/ in Serniclaes et al., 2001) and to the onset frequency of SIN2: ba975, ba1250, da1525, da1800. The initial frequency of the lowest frequency component (SIN1, the sinewave equivalent of F1) was 100 Hz and its end frequency was 750 Hz. The stimuli duration was 210 ms with a VOT of −100 ms (in French, consonants /bdg/ are prevoiced), frequency transitions of 40 ms, and a duration of the stable vocalic segment of 170 ms.

Stimuli were presented in groups of four with a SOA of 330 ms, constituting a trial. In repetition trials, one of the four stimuli was repeated four times. In other trials, a change of stimulus occurred in the fourth position. In WC trials (within-category change), the last stimulus was on the same side of the phonemic boundary as the three first (ba975 ba975 ba975 ba1250 or da1800 da1800 da1800 da1525). In AC trials (across-category change), the last stimulus was on the other side of the phonemic boundary (da1525 da1525 da1525 ba1250 or ba1250 ba1250 ba1250 da1525). The acoustical distance between the last stimulus and the preceding ones was similar in the two types of change. The only difference between AC and WC trials was that in speech mode, the change in AC trials was perceived as a change of phonemic category. Repetition, AC and WC trials were presented in a random order, and subjects had to press a key at the end of each trial to indicate whether they had perceived a change in the last stimulus, or not. The hands used to respond “same” and “different” were changed in the middle of each part and counterbalanced across subjects. Reaction times were measured from the onset of the fourth syllable with a maximum allowable response delay of 2 s. Stimuli and trial presentation, randomization, and response measurement were performed by EXPE software package (Pallier et al., 1997).

ERP experiment

Procedure

To the three critical types of trials described above (AC, WC, and repetition), we added distractor trials, in which the change between the first three syllables and the last one consisted of two steps on the continuum (e.g., ba975 to da1525). The experiment was divided into three parts comprising 300 trials each. During part 1, the subjects watched a silent movie and were told not to pay attention to the stimuli (passive part). Results from this part are not presented here. Then, during parts 2 and 3, subjects had to indicate by pressing a key whether they had or had not perceived a change in the last stimulus. In part 2, the stimuli were presented as electronic whistles (non-speech mode). After completion of this part, subjects were asked whether they had perceived these stimuli as speech sounds. In part 3, the speech-like nature of the stimuli was explained to subjects (speech mode). To obtain a switch in perception, the two extrema of the continuum were identified as /ba/ and /da/, and were then presented alternatively until the subjects heard the stimuli as syllables.

ERP recording

ERPs were collected using a 129-channel geodesic electrode net (EGI™) referenced to the vertex. This device consists in 129 Ag/AgCl electrodes encased in sponges moistened with a salty solution. The net was applied in anatomical reference to the vertex and the cantho-meatal line. Vertical eye movements and blinks were monitored via two frontal and two infra-orbital electrodes and two canthal electrodes were used to check for horizontal eye movements. Scalp voltages were recorded during the entire experiment. They were first amplified, filtered between 0.1 and 40 Hz and digitized at 125 Hz. Subsequently, the EEG was segmented into epochs starting 200 ms before the trial onset and ending 1800 ms after it. These epochs were automatically edited to reject trials contaminated by significant eye movements (deviation higher than 70 μV on the horizontal and vertical para-ocular electrodes), or body movements (local deviation higher than 70 μV and global deviation higher than 100 μV). The artifact-free trials were averaged for each subject and for each type of trials. Because ERP are very sensitive to the acoustical characteristics of the stimuli, only trials in which the 4th stimulus was the same in the three experimental conditions (that is ba1250 et da1525), were considered. The repetition trials with the two other stimuli (ba975 and da1800) and the distractor trials were discarded. The averages obtained for ba1250 and da1525 were averaged together in order to obtain three experimental conditions (repetition, AC, and WC) in each mode of perception. Averages were baseline corrected, transformed into reference-independent values using the average reference method, and digitally filtered between 0.5 and 20 Hz. Two-dimensional reconstructions of scalp voltage at each time step were computed using a spherical spline interpolation.

ERP statistical analyses

Event-related responses can be characterized as electric dipoles, revealed by positivity and negativity on the scalp. For statistical analyses, groups of electrodes were chosen to cover the negativity and the positivity maxima observed in a 2D reconstruction of the difference between conditions considered two by two. Voltage recorded at these locations and at the symmetrical locations over the controlateral hemisphere was averaged for a given time window and entered in an analysis of variance (ANOVA) with Electrodes (clusters at the positive maximum and at the negative maximum), Hemisphere (left and right), and Mode of perception (speech and non-speech) as within-subject factors. Main effects were not interpretable because of the expected voltage inversion between the selected electrodes. Therefore, only interactions between Electrodes, or Hemisphere, and the other factors of the analyses were examined.

fMRI experiment

Procedure

We used a fast-event-related paradigm similar to the ERP paradigm. Four types of trials were used: repetition, WC, AC, and silence. Each imaging sequence lasted 3.3 s starting with 300 ms of silence, followed by the four stimuli (total duration = 1220 ms), then a silence of 480 ms and finally 1300 ms of fMRI acquisition. Thus, the stimuli were presented in silence, the scanner noise being present only during the last 1300 ms of the acquisition. The experiment was divided in two parts (4 blocks of 80 trials in each part) in which the stimuli were first presented as electronic whistles then as syllables following the same procedure as in the ERP experiment.

Image acquisition and analysis

The experiment was performed on a 3-T whole body system (Bruker, Germany), equipped with a quadrature birdcage radio frequency (RF) coil and a head only gradient coil insert designed for echoplanar imaging. Functional images comprising 22 axial slices covering most of the brain were obtained with a T2-weighted gradient echo, EPI sequence (Time acquisition: 1.3 s, TR = 3.3 s, TE = 30 ms, voxel size: 3.75 × 3.75 × 5 mm). A high-resolution (1 × 1 × 1.2 mm) anatomical image using a 3-D gradient–echo inversion–recovery sequence was also acquired for each participant.

fMRI data analysis was performed using Statistical Parametric Mapping (SPM99, http://www.fil.ion.ucl.ac.uk/spm/). Preprocessing involved slice timing, movement correction, spatial normalization, and smoothing (kernel = 5 mm). The resulting functional images had cubic voxels of 4 × 4 × 4 mm^3. For each subject, a linear model was generated by entering, for each bloc, four distinct variables corresponding to the four types of trials: AC, WC, repetition, and silence. The onsets were aligned on the 4th stimuli in the trials, where a change occurred in the AC and WC conditions. The blocks were defined as non-speech or speech depending on whether subjects reported hearing syllables or whistles. However, when the subjects' response was not congruent with performance (e.g., one block with random behavioral responses in the middle of blocks with clear unambiguous responses), the block was judged ambiguous and was not included in the analyses (8/152 = 5%). The variables convoluted by the hemodynamic response function (HRF) and their temporal derivatives were included in the model. For random-effect group analyses, the individual contrast images were smoothed with a 8-mm Gaussian kernel and submitted to one-sample t tests.

To identify the circuits differentially activated in the speech and non-speech modes, the group analyses were performed within a mask comprising voxels significantly more activated in the trials with auditory stimuli (AC + WC + Rep) than in the silent trials across the entire experiment ($P < 0.001$ voxel based and $P < 0.05$ corrected for cluster extent). This mask had a size of 4657 voxels (300 cm^3) and comprised a large fronto-temporo-parietal area surrounding the Sylvian scissures, the motor cortex, the basal ganglia and cerebellum. We report the effects which pass a threshold of $P = 0.001$ for main effects and $P = 0.01$ for the interactions uncorrected for multiple comparisons and a spatial extension of $P = 0.05$ corrected for multiple comparisons. The coordinates are presented in standard Talairach and Tournoux stereotactic space (Talairach and Tournoux, 1988) and refer to maxima of the Z value within each focus of activation.

Results

Behavior

In the ERP experiment, none of the 12 subjects heard syllables before being notified. This was not the case in the fMRI experiment. Many subjects spontaneously switched to a speech mode and perceived the stimuli as syllables before we notified them. At the end of the first part, only 6 (out of 19) still heard the stimuli as non-speech sounds, 8 subjects reporting having perceived syllables as early on as the second block, i.e., after 80 trials. Self-reports were congruent with discrimination performance. Switching to the speech mode induced a decrease in errors in repetition trials and an increase in detection of a change in AC trials relative to WC trials (Fig. 1). This difference in behavior between both experiments is probably due to the degraded listening

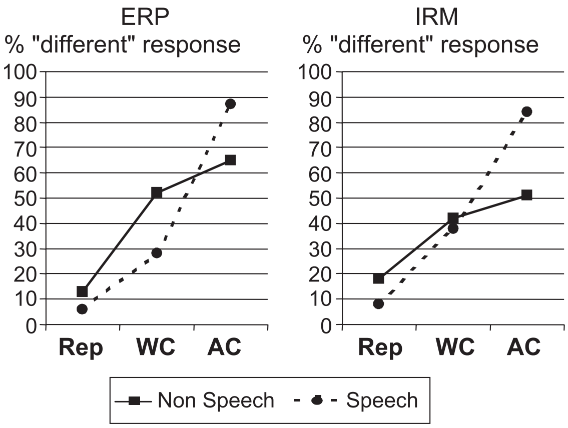

Fig. 1. Percentage of "different" responses in ERP (left) and fMRI (right) experiments.

conditions in the magnet as compared to the quiet environment of an ERP experiment.

An ANOVA computed on the percentages of "different" response in each condition using mode (non-speech vs. speech defined on the subjects' report) and condition (repetition, WC, and AC) as within-subject factors gave similar results for both experiments. As expected, the mode by condition interaction was significant [ERP: $F(2,22) = 18.3$, $P < 0.001$, fMRI: $F(2,36) = 33.9$, $P < 0.001$]. When subjects switched to the speech mode, they perceived the change in AC trials more accurately [ERP: $F(1,11) = 14.2$, $P = 0.003$, fMRI: $F(1,18) = 36.9$, $P < 0.001$], and decreased their error rate in the repetition trials [ERP: $F(1,11) = 3.7$, $P = 0.08$, fMRI: $F(1,18) = 8.4$, $P = 0.01$]. Their "different" responses in WC trials in ERPs also decreased [$F(1,11) = 14.1$, $P = 0.003$] whereas their performances in the fMRI experiment were not significantly affected by the change of mode in this condition. However, in both experiments, the difference between the detection of a change in AC and in WC trials (which defines categorical perception) increased significantly when subjects switched to a speech mode [ERP: AC–WC = 13% in non-speech mode vs. 59%, in speech mode $F(1,11) = 49.7$, $P < 0.001$, fMRI: 9% vs. 46%, $F(1,18) = 41.6$, $P < 0.001$]. Note that even in the non-speech mode, subjects perceived the changes slighter better on AC trials than on WC trials [ERP: 65% of correct detection of a change in AC vs. 52% in WC, $F(1,11) = 5.7$, $P = 0.037$, fMRI: 51% vs. 42%, $F(1,18) = 5.0$, $P = 0.038$].

In order to analyze whether there was a continuous change in performances along the experiment or an abrupt switch when subjects shifted to a speech mode, we subdivided each part of the ERP experiment in two. (This was not possible for the fMRI experiment because many subjects switched to a speech mode spontaneously after the first block of 80 trials.) The differences in percentage of a detection of change between AC and WC were 12.2% and 14.9% in the non-speech mode vs. 51.9% and 61.4% in the speech mode, demonstrating that although a small improvement (although not significant, ps > 0.1) was found between blocks in each part, a major difference occurs between the speech and non-speech modes.

Reaction times were analyzed similarly. The subjects were equally fast for both modes of perception [ERP: 844 vs. 902 ms, $F(1,11) = 1.80$, $P = 0.21$, fMRI: 757 vs. 739 ms, $F(1,18) < 1$]. However, in the speech mode, they were faster for the repetition and AC trials than for the WC trials [ERP: $F(2,22) = 4.86$, $P = 0.018$, fMRI: $F(2,36) = 21.6$, $P < 0.001$) whereas their reaction times were similar for the three conditions in the non-speech mode [ERP: $F(2,22) < 1$, fMRI: $F(2,36) = 1.3$, $P = 0.29$], yielding a significant Mode by Condition interaction [ERP: $F(2,22) = 4.3$, $P = 0.027$, fMRI: $F(2,22) = 9.8$, $P < 0.001$].

To summarize the behavioral results, switching perception from a non-speech mode to a speech mode significantly modified subjects' performances, increasing accuracy and speed for repetition and AC trials. Perception became unequivocally categorical in the speech mode although a small category effect was already present in non-speech mode (Fig. 1).

ERP

Global effect of the mode of perception

The latencies of the classical auditory responses were similar in the speech and non-speech modes: P1 56 ms, P2 154 ms, N2 218 ms, P3 312 ms, and P4 434 ms. The ERP amplitude was in general larger and with a more symmetrical distribution in the non-speech mode than in the speech mode. Significant differences between the two modes of perception were observed for each peak and for each of the four stimuli. Although the most noticeable difference was recorded at 296 ms (256–320 ms) after the onset of the first syllable of the trials (P3), a difference was present as soon as P1 (36–76 ms), consistent with a top-down attentional change. The topography of these differences was stable and predominant over the left hemisphere (Fig. 2).

To exemplify this global difference between the two modes of perception, we present two ANOVA analyses centered on P1 and P3. Voltage was averaged over a 40-ms time window centered on P1 and P3 after each syllable at four locations on each hemisphere (cluster of five electrodes at fronto-central, infero-frontal, temporal, and occipital location). ANOVAs were computed with Stimulus number (1 to 4), Electrodes location, Hemisphere, and Mode (non-speech and speech) as within-subject variables. The difference between the speech and non-speech modes was demonstrated by the significant Electrode by Mode interaction present for P1 [$F(3,33) = 3.23$, $P = 0.035$] and for P3 [$F(3,33) = 12.48$, $P < 0.0001$]. The topography of the difference tended to be asymmetric, being present mainly over the left hemisphere. This asymmetry was evident when the analysis was restricted to the most noticeable difference [i.e., over the P3 after the first syllable, Electrode × Mode × Hemisphere: $F(3,33) = 4.22$, $P = 0.01$], but only marginal when ANOVAS were computed over the entire trial [for P3: $F(3,33) = 2.56$, $P = 0.07$ and P1: $F(3,33) = 2.34$, $P = 0.09$], certainly because of a floor effect due to the decreasing amplitude of ERP responses to repeated stimuli. Similar results were obtained for the other peaks.

Mismatch responses

The responses to the stimulus change were analyzed from the onset of the 4th stimulus on. We report and analyze only the early differences, or mismatch responses, between conditions. To determine the topography and the latency of the mismatch responses, we have inspected the time course of two-dimensional reconstructions of the differences (AC–repetition) and (WC–repetition) in both modes. In the non-speech mode, a bilateral frontal negativity developed between 150 and 190 ms post onset of the 4th stimulus, in synchrony with a right temporal positivity for both contrasts. In the speech mode, a right frontal negativity synchronous with a left temporal positivity began earlier at 100 ms, developed until 190 ms and slowly vanished until 270 ms for the AC–repetition contrast. For the WC–repetition contrast, a smaller mismatch response, but with a similar topography to that for the AC contrast, was present from 100 to 150 ms. This response vanished and reappeared from 210 to 274 ms, with a median frontal negativity synchronous with bilateral temporal positivities (Figs. 3 and 4).

To analyze these mismatch responses, four groups of electrodes (clusters of six to eight electrodes at the antero-frontal, latero-frontal, fronto-central, and temporal locations) were chosen according to the topography of the mismatch responses in the different contrasts, and analyses of variance were performed on three contrasts (AC vs. Repetition, WC vs. Repetition, and AC vs. WC) and across three time windows that encompassed the different mismatch responses (100–150, 150–190, 210–274 ms).

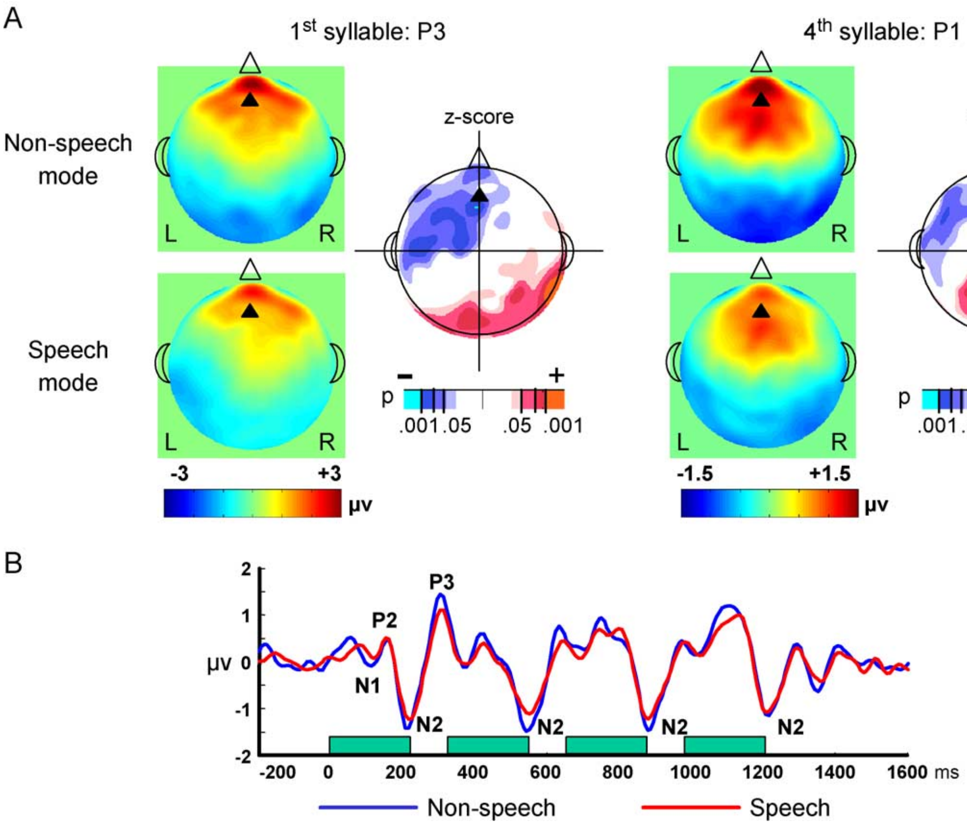

Fig. 2. (A) Two examples of the global differences between speech and non-speech mode in ERP recordings. Left voltage topography and *z*-score maps at 304 ms (P3) after the onset of the first syllable, and right at 56 ms (P1) after the onset of the 4th syllable. (B) Grand average recorded from Fz in speech and non-speech mode. The green rectangles correspond to the duration of the stimuli.

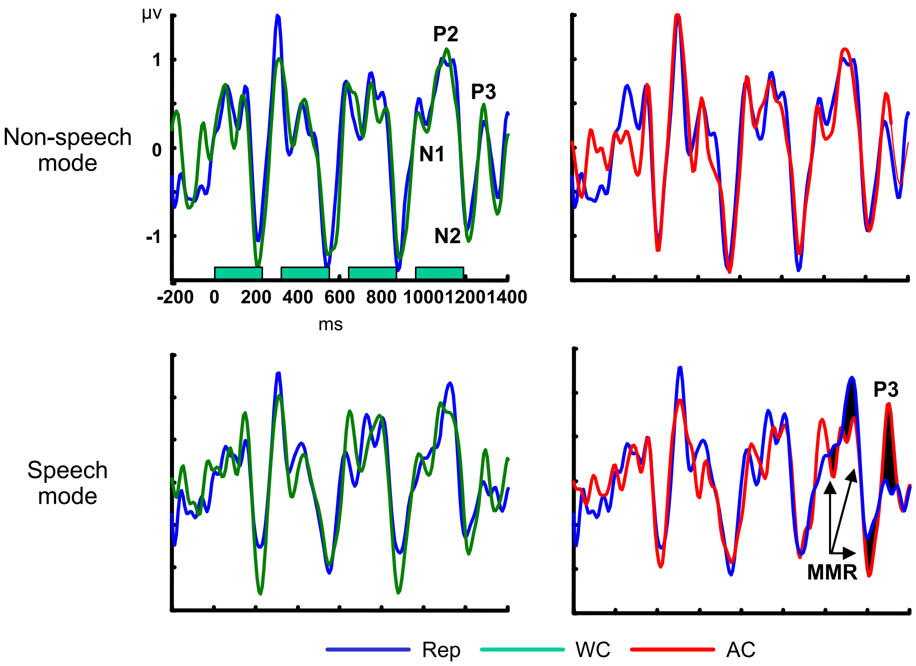

Fig. 3. Grand average recorded from Fz for the three conditions in speech and non-speech mode. For this electrode, the MMR is particularly evident for the (AC–Rep) contrast in speech mode. It extends over N1, P2, and N2.

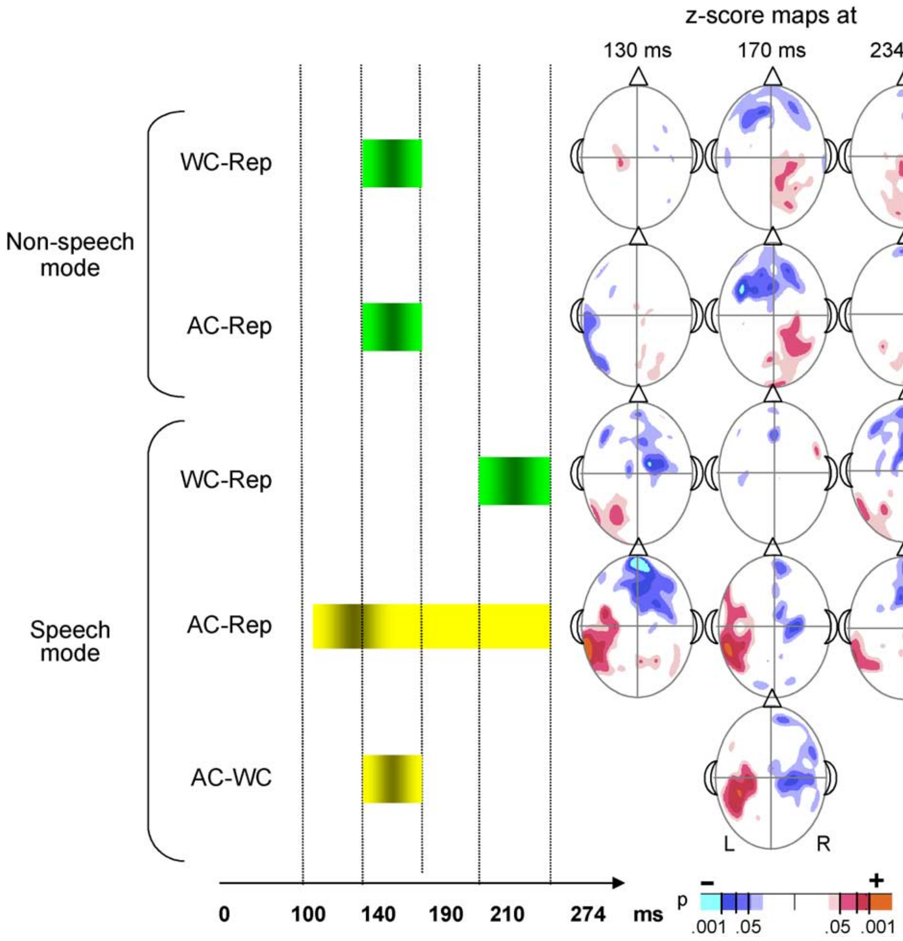

Fig. 4. ERP mismatch responses after the 4th stimulus for the different conditions in speech and non-speech mode. Left, the colored rectangle identifies the time window during which an MMR was present. Green rectangles correspond to comparisons isolating an acoustical effect and yellow to comparisons isolating a phonemic effect. Right, MMR topographies (*z*-score maps) at the maximum of the response.

Non-speech mode

For the first time window (100–150 ms), no mismatch response was present in the non-speech mode [Electrode × Condition: $F(3,33) < 1$ for all contrasts]. The mismatch response appeared in the second time window (150–190 ms), and was similar for both changes [Electrode × Condition for WC–Repetition: $F(3,33) = 3.1$, $P = 0.041$, AC–Repetition: $F(3,33) = 4.09$, $P = 0.014$ and AC–WC: $F(3,33) < 1$]. In the third time window, the mismatch response was extinct (ps > 0.15). No interaction with hemisphere was significant in these analyses.

Speech mode

For the first time window (100–150 ms), a mismatch response was recorded for the phonemic contrast AC–Repetition [Electrode × Condition: $F(3,33) = 4.9$, $P = 0.006$] but not for the acoustic contrast WC–Repetition [$F(3,33) < 1$]. However, the difference between AC and WC did not reach significance [Electrode × AC–WC: $F(3,33) = 2.2$, $P = 0.11$]. Post hoc analyses of the phonemic mismatch response restricted to each electrode location revealed significant asymmetries over the temporal and the frontocentral channels [Hemisphere × AC–Rep, respectively: $F(1,11) = 6.9$, $P = 0.023$ and $F(1,11) = 5.2$, $P = 0.043$].

For the second time window (150–190 ms), the mismatch response decreased for the phonemic contrast but was still present [Electrode × Condition: $F(3,33) = 2.75$, $P = 0.059$]. No mismatch response was present for the acoustic contrast WC–Repetition and a significant interaction Hemisphere × Electrode × AC–WC was observed [$F(3,33) = 3.5$, $P = 0.025$]. Post hoc analyses of the phonemic mismatch response showed the same significant asymmetries over the temporal and the frontocentral channels as for the previous window for the contrast AC–Repetition (Hemisphere × Condition: ps < 0.02) but also for the contrast AC–WC (ps < 0.031).

For the third time window (210–274 ms), the phonemic mismatch response was still present, although weak [Electrode × Condition $F(3,33) = 2.8$, $P = 0.055$] and for the first time a significant mismatch response was observed for the acoustic contrast WC–Rep [$F(3,33) = 3.1$, $P = 0.038$]. No significant difference between the two mismatch responses was present. No significant interaction with hemisphere was present during this time window.

Interactions with mode of perception

If the time course of the mismatch responses is affected by mode of perception, we expect a significant interaction Electrodes × Mode × AC–Rep during the first time window. Indeed, it is what was obtained [$F(3,33) = 2.8$, $P = 0.054$]. This was due to an effect of mode for the AC condition [Electrode × Mode: $F(3,33) = 3.9$, $P =$

0.016]. No such interaction was obtained for the acoustic contrast WC–Rep (WC–Rep × Electrodes × Mode: $F < 1$).

During the second time window, a mismatch response developed for the acoustic contrasts in the non-speech mode, while the acoustic contrast in the speech mode (WC–Rep) was not significant. This created a marginally significant interaction WC–Rep × Electrodes × Mode [$F(3,33) = 2.9$, $P = 0.051$]. No interaction with mode of perception was present during the third time window.

Although the mismatch response topographies appear to be differently distributed over the left and right hemispheres in the speech and non-speech modes, especially over the temporal areas (Fig. 4), no significant effect of hemisphere was found in the ANOVAs.

fMRI

Regions affected by mode of perception

The first analysis aimed at identifying the circuits responding differentially when the subject perceived the stimuli as syllables or as whistles, by looking at the interaction between Mode (speech–non-speech) and the contrast Sound (AC + WC + Rep) minus Silence. The only region that appeared was the posterior part of the left superior temporal gyrus extending along the superior temporal sulcus (see Table 1, and Fig. 5). We checked that this cluster was identified in the interactions between Mode (speech–non-speech) and each of the three conditions minus Silence. The cluster extended over 57 voxels (z value = 3.96 at local maximum) for AC by Mode, 30 voxels (z = 2.96) for WC by Mode, and 80 voxels (z = 3.07) for Rep by Mode. The reverse interaction (non-speech − speech) × (sound − silence) did not identify any areas that were more active in the non-speech than in the speech mode.

Table 1
Brain areas more activated in speech mode than in non-speech mode

Area	No. of voxels in cluster	Cluster-level P value (corrected)	Z value at local maximum	Talairach coordinates		
				x	y	z
1. Sound–Silence × Speech–Non-speech						
Left post STS (area 22)	69	0.001	3.09	−56	−40	0
Left post STG (area 22)				−60	−24	4
2. AC–Rep × Speech–Non-speech						
Left SMG (area 40)	27	0.019	3.34	−52	−52	44
Right SMG (area 40)	27	0.019	2.77	40	−44	52
Left thalamus	19	0.069	3.27	−12	−16	0
3. AC–WC × Speech–Non-speech						
Left SMG (area 40)	43	0.004	2.97	−52	−36	32
			2.96	−48	−52	40
4. WC–Rep × Speech–Non-speech						
No activated areas						

STS: superior temporal sulcus; STG: superior temporal gyrus; SMG: supramarginal gyrus.

The second series of analyses aimed at identifying the networks involved in acoustic and phonemic discrimination (Table 2). In the non-speech mode, all the changes were by definition acoustic: The areas significantly more activated by a change of stimulus than by repetition (WC+AC–Rep) were left and right subcortical structures (thalami, putamen, caudate), and a left inferior frontal cluster extending from the anterior insula to the frontal operculum. No significant differences were present between the two types of change, (AC–WC) and (WC–AC). In speech mode, the (WC–Rep) contrast is also an acoustic change with no linguistic value. This contrast identified regions similar to those in non-speech mode: left thalamus and putamen, left insula extending to the frontal operculum, and a cluster in the left anterior cingulate gyrus. The WC–Rep contrast yielded similar activations in speech mode and in non-speech mode.

The circuit involved in the detection of a phonemic change in the speech mode (AC–Rep) involved left thalamus and basal ganglia, both insula, the anterior cingulate gyrus, the posterior part of the left superior temporal sulcus and the infero-parietal regions on both sides (Fig. 5). The regions in that network that are specific to a phonemic network are revealed by the AC–WC contrast in the speech mode and by the interactions of AC–Rep and AC–WC with mode of perception. The first comparison isolated the left and right supramarginal gyri that were more activated in AC than in WC in speech mode. The second comparison (Mode of perception × AC–Rep) found the same regions plus the left thalamus while Mode of perception × AC–WC isolated only the left supramarginal gyrus. Thus, the supramarginal gyri, especially on the left, were significantly more activated by a phonemic change than by acoustic changes. No areas were isolated in the (WC–AC), (Rep–AC), and (Rep–WC) contrasts.

To examine decision and response processes, we performed an analysis contrasting trials in which the subjects respond “same” vs. trials in which they respond “different”. This analysis identified the left thalamus and basal ganglia, the anterior cingulate, the left insula, and the left operculum. No significant cluster was identified by the interaction of this contrast with mode of perception.

Lateralization

To study whether these networks were significantly lateralized, smoothed contrast images were flipped by applying a linear transformation matrix that switched the left and the right hemispheres for each subject. We then obtained individual asymmetry images by subtracting the original contrast image from their flipped versions. Those images were entered into a random-effect analysis, which tested whether the amount of activation for the contrast of interest was significantly larger in one hemisphere relative to the other (voxel $P < 0.001$, cluster $P < 0.05$ corrected). In both modes, all the areas activated by the stimuli (sound–silence contrast) were significantly asymmetric favoring the left side. Only a cluster in the left supramarginal gyrus was more asymmetric in the speech mode than in the non-speech mode (Table 3).

Discussion

Because of their spatial and temporal characteristics, ERP and fMRI do not provide the same types of information. Therefore, we will first discuss the ERPs results in order to study how the mode of perception affects the first stages of processing of the auditory

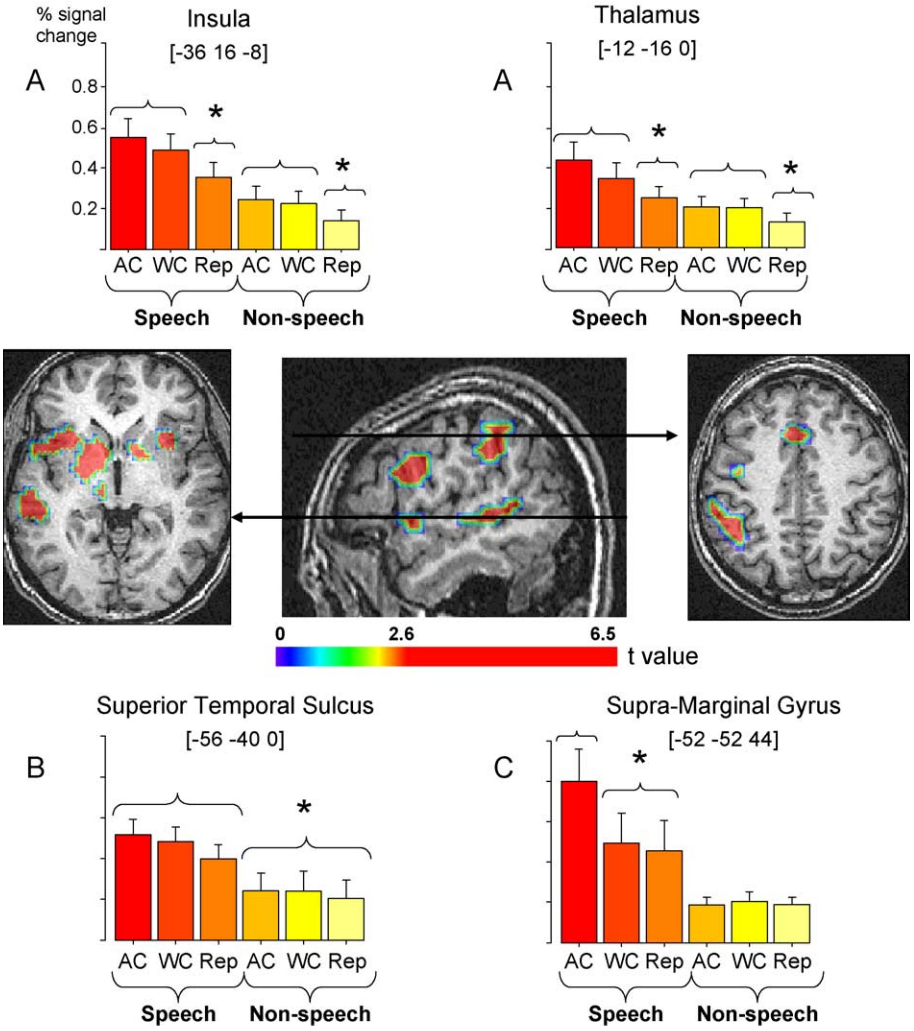

Fig. 5. AC minus Rep in speech mode. Sagittal and axial maps showing significant clusters of the random analysis superimposed on a representative brain. This contrast illustrates the three functional regions isolated in the discrimination task: (A) anterior and subcortical regions sensitive to a change of stimulus, (B) posterior part of the superior temporal sulcus sensitive to the mode of perception, and (C) supramarginal gyrus sensitive to a change of phoneme. See Table 2 for the full description of the extension and localization of the activated clusters in this contrast. Bar diagrams indicate the percentage of signal change from silence for the three conditions in each mode at the indicated locations. A star signals a significant comparison between the conditions in brackets.

signal. Then, the fMRI results will help to determine the localization of the involved networks.

ERPs

Our paradigm allows a comparison of the ERPs to the same stimulus presented in different context conditions. If a significant difference, or mismatch response, is present, it can only be due to the computation of the difference between the target and the preceding context. This response gives us access to the code of the auditory representations that are stored in sensory memory. Our goal was thus to study how the properties of the mismatch response (latency, duration, and topography) are affected by the mode of perception (speech vs. non-speech) and by the nature of the change in speech mode (WC: acoustic vs. AC: phonemic).

The present data allow rejection of several hypotheses. First, the detection of a change between two auditory stimuli could have depended only on the physical distance between the stimuli. Because the stimuli are physically identical in both modes of perception, the mismatch response (MMR) should have been similar in the speech and non-speech mode and in acoustic and phonemic trials. This was not the case. For the phonemic contrast (AC condition in speech mode), the onset of the MMR was fast and the MMR extended over 100 ms whereas in the acoustic contrasts, the onset of a significant MMR was delayed and the MMR duration shorter. In particular, the MMR for an identical change (AC–Rep) was significantly different in the two listening modes suggesting that a more efficient network was involved when the stimuli were processed as speech in the second half of the experiment.

A second hypothesis may explain the efficiency of the discrimination network in the speech mode not because of the involvement of a different network in this mode (e.g., phonemic vs. acoustic) but because of longer exposure to the stimuli when subjects are in the second part of the experiment. After long exposure to the same stimuli, subjects would have learned to perceive the slight differences between the stimuli. A first argument against this hypothesis is the abrupt change in performance when subjects shifted to speech mode. Second, learning would imply an improvement of the discrimination responses in all

Table 2
Brain areas activated by the detection of a change

Area	No. of voxels in cluster	Cluster-level *P* value (corrected)	*Z* value at local maximum	Talairach coordinates		
				x	*y*	*z*
1. Acoustical change in non-speech mode (AC + WC–Rep)						
Right thalamus/basal ganglia	132	<0.001	4.10	8	0	12
Left thalamus/basal ganglia			3.60	−16	8	−16
Left anterior insula	44	0.001	2.87	−36	16	−4
Left infero frontal gyrus			2.67	−56	8	4
Left precentral gyrus	31	0.003	3.48	28	−16	40
2. Acoustical change in speech mode (WC–Rep)						
Left anterior insula	212	<0.001	4.04	−32	24	−4
Left infero frontal gyrus			3.88	−52	12	0
Left precentral gyrus			3.63	−56	4	32
Left cingulate gyrus	69	<0.001	3.71	−8	20	40
Right anterior insula	32	0.013	3.67	36	16	−4
Left thalamus/basal ganglia	63	<0.001	3.66	−12	−8	4
3. Phonetic change in speech mode (AC–Rep)						
Right insula	470	<0.001	4.62	32	16	−8
Left anterior insula			4.20	−36	16	−8
Left thalamus/basal ganglia			4.07	−8	12	4
Left precentral sulcus	56	<0.001	3.94	−32	12	48
Left SMG	115	<0.001	3.88	−52	−52	40
Left cingulate gyrus	78	<0.001	3.65	−12	8	52
Right SMG	111	<0.001	3.63	48	−32	52
Left posterior STG and STS	40	0.001	3.34	−60	−36	−4

conditions. This was not observed. Whereas responses were significantly faster and more accurate for repetition and AC trials in the speech than in the non-speech mode, there was no effect of mode on reaction times for WC (918 vs. 923 ms) and accuracy in the detection of a change in this condition decreased (40% vs. 22%). Furthermore, the MMR was delayed for the acoustic contrast in speech mode, becoming significant only after 210 ms; that is 60 ms after the onset of the MMR for the same contrast in the non-speech mode. This suggests that the differences observed between the speech and non-speech modes were not related to duration of exposure but rather to a switch in perception involving partially distinct neural processes.

Our results are thus best explained by a third hypothesis stating that the discrimination responses for AC do not involve the same networks in both modes. A phonemic representation is computed in speech mode as demonstrated by the categorical effect for the MMRs in the speech mode (i.e., a significant difference between MMRs for AC and WC), whereas no difference was observed

Table 3
Left–right asymmetries (left > right)

Area	No. of voxels in cluster	Cluster-level *P* value (corrected)	*Z* value at local maximum	Talairach coordinates		
				x	*y*	*z*
1. Sound–Silence × Speech–Non-speech						
Left thalamus	24	0.006	3.21	−20	−20	4
Left SMG	17	0.031	2.76	−32	−64	48
2. AC–Rep × Speech–Non-speech						
Left SMG	16	0.047	3.27	−52	−52	44
3. AC–WC × Speech–Non-speech						
Left SMG	32	0.001	3.12	−52	−48	40
Left anterior insula	20	0.016	3.94	−44	12	−8
4. WC–Rep × Speech–Non-speech						
No activated areas						
5. WC–Rep × Non-speech– Speech						
Left anterior insula	15	0.051	3.08	−40	16	−4

between these two conditions in the non-speech mode. Several ERP experiments have already demonstrated that mismatch responses are sensitive to properties of phoneme perception such as categorical perception, normalization, and subjects' native language. As here, the MMR for a nonpertinent linguistic change (i.e., within-category changes or changes that cross non-native phonemic boundaries) is generally weaker, or sometimes absent, compared to the robust phonemic MMR (Aaltonen et al., 1992; Dehaene-Lambertz, 1997; Näätänen et al., 1997; Phillips et al., 2000; Rivera-Gaxiola et al., 2000; Sharma and Dorman, 2000; Winkler et al., 1999). Here, we add that the categorical effect observed for linguistic MMR cannot be explained by an underlying discontinuity in acoustic mismatch responses.

Previous work has shown that a longer duration and an earlier onset of mismatch responses are correlated with better discrimination performances (Tiitinen et al., 1994; Tremblay et al., 1998). This is confirmed in the present experiment. The shorter latency and faster duration of the MMR, associated with better accuracy in performance and faster reaction times, suggest that the phonemic representation is computed faster and more efficiently than the acoustic representation of the same stimuli. Furthermore, because of the slower MMR for WC in speech mode as compared to the non-speech mode (60-ms delay), it is possible that the phonemic network, once activated by speech, exerts an inhibitory influence on the concurrent auditory representations to prevent interference from nonlinguistically pertinent differences (Liberman et al., 1981; Liebenthal et al., 2003).

Phonemic perception has been strongly associated in the literature with the left hemisphere. Here, a global difference between the speech and non-speech modes of perception was observed mainly over the left hemisphere. Furthermore, the phonemic MMR was more lateralized than the acoustic MMRs for which no interaction with hemisphere was significant. A similar topography has been already described for phonemic discrimination (Dehaene-Lambertz, 1997; Dehaene-Lambertz et al., 2000; Näätänen et al., 1997). Näätänen et al. (1997), using MEG, localized the origin of the phonemic response to the left planum temporale. Because of the difficulties of localizing active regions with ERPs, we will now examine the results obtained with fMRI.

fMRI

To summarize the fMRI results, we have isolated three sets of regions with distinct response profiles mode (Fig. 5). The first one located in the posterior part of the temporal lobe along the left superior temporal sulcus, is more activated in all speech conditions relative to non-speech conditions. The second in the left supramarginal gyrus is significantly more asymmetric in the speech than in the non-speech mode and is sensitive to the presence of a linguistic change. This region is not activated during trials with an acoustic change even when the stimuli are physically identical (e.g., AC–Rep contrast in non-speech mode). The third region comprised the thalami, the basal ganglia, the insula and the frontal operculum, mainly in the left hemisphere and is more activated in trials with a change of stimulus relative to repetition trials or in trials in which subjects respond "different" relative to the trials in which they respond "same".

Neuropsychological studies have reported the left superior temporal and supramarginal regions as being the more common affected areas in phoneme processing deficits (Caplan et al., 1995; Dronkers et al., 2000). Neuroimaging studies have also found both regions to be activated in linguistic perceptual tasks (Binder et al., 2000; Burton et al., 2000; Celsis et al., 1999; Perani et al., 1996). In our experiment, the posterior temporal region was more sensitive to the mere property of the stimuli being perceived as speech independently of the condition. In each mode, the percentage of signal change was almost similar in all types of trials (AC, WC, and repetition). The supramarginal activation was more specific, being significantly present only in trials with a phonemic change. This suggests that these two regions may play different roles. The involvement of the temporal cortex in the speech non-speech contrast is congruent with several studies showing that the superior temporal region responds more strongly to speech than to non-speech sounds (Belin et al., 2002; Benson et al., 2001; Binder et al., 2000; Jäncke et al., 2002; Perani et al., 1996; Vouloumanos et al., 2001). In a meta-analysis of four studies, Binder et al. (2000) reported that the mean of the peaks of significant activation differences for speech and non-speech sounds is at $x = -55.5$ (SD 2.3), $y = -20.2$ (SD 10.9), $z = 0.3$ (SD 4.1) in the left hemisphere. This point lies within the anterior part of our cluster, which also extends backward to the posterior superior temporal sulcus. The latter appears essential in syllable processing. Boatman et al. (1997) have reported in five patients that the stimulation of one temporal electrode in this region elicited a deficit in syllable discrimination. The more posterior part of the temporal region is also activated when phonological representations are accessed through devices other than auditory input, such as lip reading (Calvert and Campbell, 2003; Calvert et al., 1997), word generation (Buchsbaum et al., 2001; Wise et al., 2001), sign language processing (Petitto et al., 2000), and reading (Paulesu et al., 2000, 2001). It is possible that top-down attention amplifies those speech representation, thus bringing knowledge of the characteristics of a human voice to supplement the poverty of the sinewave input and integrate its acoustical features into phonemic representations. Although sinewave speech does not sound like human speech, the peaks of activation to vocal productions reported by Belin et al. (2000) are located at the anterior and internal part of our cluster (e.g., $x = -60$, $y = -23$, $z = 6$ in experiment 3; Belin et al., 2000). An effect of mode of perception is detected as soon as 56 ms (P1) in the ERPs. The origin of P1 has been attributed to activity in the superior temporal gyrus by epidural recordings (Eggermont and Ponton, 2002). This region, defined as the parabelt in humans, is connected to the upper and lower banks of the superior temporal sulcus. The electrophysiological result is thus compatible with an early activation of the temporal fMRI cluster. Given the background literature, we tentatively propose that the superior temporal activations that we observed reflect the greater deployment of early speech-specific phoneme and syllable processing in speech mode.

The fact that this temporal cluster is not sensitive to a change of phoneme in the present study is unexpected but may be related to two factors. First, our paradigm is based on a habituation–dishabituation design and assumes that repetition induces a decrease of activity that recovers when a change of stimulus is perceived. It is possible that no habituation occurs in this region because in speech, repetition carries as much as information as a change. For example, /dada/ and /daba/ are two possible words. Thus, each stimulus whether standard or deviant, elicits the same activation and no difference is detected when a change of stimulus occurs. However, repetition induced a decrease in ERPs amplitude even when the sinewave stimuli were perceived as syllables, thus contradicting this hypothesis. A second explanation relates to the difficulties of the task. The stimuli were impoverished syllables that differed subtly, and were presented in the noisy environment of fMRI. Although the stimuli were presented during a short silent period of the scanner, the magnet noise could have had a forward or backward masking effect. Attentional effects or other top-down processes might thus have activated this region more in fMRI relative to what was recorded during the quiet and less stressful ERP experiment. This region may also be reactivated at the end of each trial if subjects check whether their response was correct. In both cases, this region would be activated similarly for repetition trials and change trials because fMRI, contrary to ERPs, cannot separate early and late events occurring during perception.

We now turn to the cluster sensitive to phonemic changes, located in the supramarginal gyrus. Although activations in the left inferior parietal region are commonly found in neuroimaging studies using syllable discrimination tasks (Burton et al., 2000; Celsis et al., 1999; Jacquemot et al., 2003), it is not clear whether its role is purely sensory or more indirect, for instance reflecting retrieval in verbal working memory (Paulesu et al., 1993). In our experiment, the subjects' performance was clearly different in AC trials in speech mode relative to the other conditions. Our experiment thus cannot easily disentangle early encoding processes from later decision or memory processes.

Several models have tried to explain the role of the inferior parietal cortex in phoneme processing. In primates, the cortical auditory system has been divided into two processing streams, a spatial stream projecting to the parietal cortex from the superior temporal gyrus and an object stream associated with vocalizations projecting to the anterior temporal and frontal areas (Poremba et al., 2003; Rauscheker and Tian, 2000). A similar distinction has been proposed for the human cortex. However, if the concept of multiple parallel processing streams can certainly be generalized to the human auditory system, the attribution of a "what" and a "where" role to the ventral and the parietal stream, respectively, should be clarified. The role of the human parietal lobe in phoneme processing is not easily explained by a "where" function except after a broad extension of the notion of "where". Wise et al. (2001) suggest that spatial attention is necessary to direct motor responses, among them articulatory responses especially during the repetition of speech and non-speech sounds. Belin and Zatorre (2000) suggest that spatial localization and spectral motion, important for speech perception, share many properties and could be processed by networks sharing similar features, such as a dorsal localization in the auditory system. Hickok and Poeppel (2000) propose another model based also on a dual-route concept. For them, after a sound-based representation of speech is constructed in the posterior superior temporal region, one pathway goes ventrally and is involved in the interfacing of these representations with widely distributed conceptual representations. The second pathway involves the inferior parietal and frontal system and is involved in tasks that require explicit access to sublexical speech segments, because these structures are part of a sensory–motor interface system. The fact that our parietal cluster was not activated by all changes but only by the phonemic change does not favor a general explanation in terms of spectral motion of the auditory signal, as proposed by Wise et al. (2001) or Belin and Zatorre (2000). Its localization, close to sensory representations of the mouth, rather implies a sensory–motor interface as suggested by Hickok and Poeppel (2000). Buccino et al. (2001) reported that observation of mouth movement activates a region close to our cluster ($x = 52$ $y = -32$ $z = 44$ and $x = -36$ $y = -52$ $z = 44$ in their study vs. $x = 40$ $y = -44$ $z = 52$ and $x = -52$ $y = -52$ $z = 44$ in ours). Although the movements studied by Buccino et al. were not related to speech, we can postulate that effector-related representations of speech movements are also present in the parietal lobe and may contribute to the phonological store of short-term verbal memory (Paulesu et al., 1993). In this context, the activation observed in the supramarginal gyrus might correspond to temporary phonological storage after a perceived change of syllable.

The third interesting region comprises subcortical structures, insula, and frontal operculum. Although neuropsychological and electrical stimulation studies have found that posterior sites are those mostly responsible for phonemic processing, some functional imaging studies have reported activations in Broca's area for phonemic discrimination and phoneme monitoring tasks (Burton et al., 2000; Zatorre et al., 1992, 1996). These activations have been related either to articulatory gestures in order to process fine-grained distinctions (Zatorre et al., 1996), or to segmentation of syllables into phonemic units (Burton et al., 2000). However, frontal activations in the same regions have also been observed in experiments involving non-linguistic stimuli (Fiez et al., 1995) and even visual stimuli (Linden et al., 1999), suggesting that they may be not specific of phonemic processing but related to response selection when a target is detected, or to attentional demand when stimuli are complex. In a recent paper, Binder et al. (2004), using a speech identification task in noise, were able to distinguish between a sensory component located in the superior temporal areas, correlated with accuracy of identification, and a decision component located in the anterior insula and adjacent medial frontal opercular cortex, strongly correlated with choice RT. The frontal cluster activated in our experiment in all change conditions, either acoustic or phonemic, are very closed to the frontal region identified by Binder et al. Because neither articulatory gestures, nor syllable segmentation seem relevant to explain the activation observed in non-speech mode, our result confirms Binder et al.'s interpretation of frontal activations as relating to decision process.

Hemispheric asymmetries

The lateralization of phoneme perception has been attributed by some authors to the acoustical structure of speech stimuli in conjunction with structural asymmetries between the left and right hemispheres. More fine and myelinated fibers in the left hemisphere would tend to process fine temporal information while large fibers in the right hemisphere would be more efficient to process spectral information (Zatorre and Binder, 2000). A leftward asymmetry to speech sounds is indeed observed in other mammals: For example, King et al. (1999) recorded stronger evoked responses in the left thalamus of guinea pigs in response to speech stimuli (CV syllables). Our results partially support this hypothesis. On the one hand, our stimuli were processed asymmetrically favoring the left hemisphere in both speech and non-speech mode. However, on the other hand, activation in the left thalamus and the left supramarginal gyrus was even more asymmetric in the speech than in the non-speech mode. This suggests that general auditory characteristics are not sufficient to explain hemispheric asymmetry in phoneme perception. The left hemisphere presents a specialization for speech stimuli that goes beyond stimulus-driven characteristics; its contribution is best described here as providing a top-down "speech processing" mode, even when the stimuli can be ambiguously processed as speech or non-speech.

Conclusion

The present work emphasizes three points. First, the same auditory stimuli are processed differentially depending on whether they are perceived as speech or as nonsense electronic whistles. Second, the posterior part of the superior temporal sulcus and the supramarginal gyrus are crucial areas for syllable processing but are not involved in the processing of the same physical dimension when the stimuli are not perceived as speech. Third, non-phonemic auditory representation and phonemic representation are computed in parallel, but the phonemic network is more efficient and its activation may have an inhibitory effect on the acoustical network. These properties validate the notion of a distinct speech mode in the human brain. Further research should determine to what extent this organization is laid down in the course of language acquisition and to what extent it is already present early in infancy (Dehaene-Lambertz and Baillet, 1998; Dehaene-Lambertz et al., 2002; Eimas et al., 1971).

Acknowledgments

This study was supported by IFR49, Contrat BioMed No. PSS*1046, "Imaging of Language Functions in the Brain", 1999–2001, ACI Blanche 1999 "Plasticité neuronale et acquisition du langage", CTI-53. Perception catégorielle et dyslexie: Des données comportementales à la neuro-imagerie; ACI-129. Traitement de la parole et dyslexie (Cognitique, MENRT), Progres INSERM.

References

Aaltonen, O., Paavilainen, P., Sams, M., Näätänen, R., 1992. Event-related brain potentials and discrimination of steady-state vowels within and between phonemes categories: a preliminary study. Scand. J. Log Phon. 17, 107–112.

Belin, P., Zatorre, R.J., 2000. 'What', 'where' and 'how' in auditory cortex. Nat. Neurosci. 3 (10), 965–966.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Brain Res., Cogn. Brain Res. 13 (1), 17–26.

Benson, R.R., Whalen, D.H., Richardson, M., Swainson, B., Clark, V.P., Lai, S., et al., 2001. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain Lang. 78 (3), 364–396.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P.W., Kennedy, L., Mehler, J., 1988. An investigation of young infants' perceptual representations of speech sounds. J. Exp. Psychol. Gen. 117, 21–33.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., et al., 2000. Human temporal lobe activation by speech and non speech sounds. Cereb. Cortex 10 (5), 512–528.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7 (3), 295–301.

Boatman, D., Hall, C., Goldstein, M.H., Lesser, R., Gordon, B., 1997. Neuroperceptual differences in consonant and vowel discrimination: as revealed by direct cortical electrical interference. Cortex 33, 83–98.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., et al., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur. J. Neurosci. 13 (2), 400–404.

Buchsbaum, B.R., Hickok, G., Humphries, C., 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cogn. Sci. 25, 663–678.

Burton, M.W., Small, S.L., Blumstein, S.E., 2000. The role of segmentation in phonological processing: an fMRI investigation. J. Cogn. Neurosci. 12 (4), 679–690.

Calvert, G.A., Campbell, R., 2003. Reading speech from still and moving faces: the neural substrates of visible speech. J. Cogn. Neurosci. 15 (1), 57–70.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., et al., 1997. Activation of auditory cortex during silent lipreading. Science 276 (5312), 593–596.

Caplan, D., Gow, D., Makris, N., 1995. Analysis of lesions by MRI in stroke patients with acoustic–phonetic processing deficits. Neurology 45 (2), 293–298.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I., Chollet, F., 1999. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. NeuroImage 9, 135–144.

Dehaene-Lambertz, G., 1997. Electrophysiological correlates of categorical phoneme perception in adults. NeuroReport 8 (4), 919–924.

Dehaene-Lambertz, G., Baillet, S., 1998. A phonological representation in the infant brain. NeuroReport 9, 1885–1888.

Dehaene-Lambertz, G., Dupoux, E., Gout, A., 2000. Electrophysiological correlates of phonological processing: a cross-linguistic study. J. Cogn. Neurosci. 12 (4), 635–647.

Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L., 2002. Functional neuroimaging of speech perception in infants. Science 298, 2013–2015.

Dronkers, N.F., Redfern, B.B., Knight, R.T., 2000. The neural architecture of language disorders. In: Gazzaniga, M.S. (Ed.), The New Cognitive Neurosciences, second ed. MIT Press, Cambridge, pp. 949–958.

Eggermont, J.J., Ponton, C.W., 2002. The neurophysiology of auditory perception: from single units to evoked potentials. Audiol. Neuro-Otol. 7 (2), 71–99.

Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., Vigorito, J., 1971. Speech perception in infants. Science 171, 303–306.

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

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4 (4), 131–138.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., Dupoux, E., 2003. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. J. Neurosci. 23, 9541–9546.

Jäncke, L., Wüstenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. NeuroImage 15, 733–746.

King, C., Nicol, T., McGee, T., Kraus, N., 1999. Thalamic asymmetry is related to acoustic signal complexity. Neurosci. Lett. 267 (2), 89–92.

Kluender, K.R., Greenberg, S., 1989. A specialization for speech perception? Science 244, 1530.

Kluender, K.R., Diehl, R.L., Killeen, P.R., 1987. Japanese Quail can learn phonetic categories. Science 237, 1195–1197.

Kuhl, P.K., Miller, J.D., 1975. Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants. Science 190, 69–72.

Liberman, A.M., 1996. Speech: A Special Code. Bradford Books/ MIT Press, Cambridge, MA.

Liberman, A.M., Mattingly, I.G., 1989. A specialization for speech perception. Science 243 (4890), 489–494.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychol. Rev. 74, 431–461.

Liberman, A.M., Isenberg, D., Rakerd, B., 1981. Duplex perception of cues for stop consonants: evidence for a phonetic mode. Percept. Psychophys. 30 (2), 133–143.

Liebenthal, E., Binder, J.R., Piorkowski, R.L., Remez, R.E., 2003. Short-term reorganization of auditory analysis induced by phonetic experience. J. Cogn. Neurosci. 15 (4), 549–558.

Linden, D.E.J., Prvulovic, D., Formisano, E., Völlinger, M., Zanella, F.E., Goebel, R., et al., 1999. The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. Cereb. Cortex 9, 815–823.

Näätänen, R., Lehtokovski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385, 432–434.

Pallier, C., Dupoux, E., Jeannin, X., 1997. EXPE: an expandable programming language for on-line psychological experiments. Behav. Res. Methods Instrum. Comput. 29, 322–327.

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

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., et al., 2000. A cultural effect on brain function. Nat. Neurosci. 3 (1), 91–96.

Paulesu, E., Demonet, J.F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., et al., 2001. Dyslexia: cultural diversity and biological unity. Science 291 (5511), 2165–2167.

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S.F., Dupoux, E., et al., 1996. Brain processing of native and foreign languages. NeuroReport 7, 2439–2444.

Petitto, L.A., Zatorre, R.J., Gauna, K., Nikelski, E.J., Dostie, D., Evans, A.C., 2000. Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. Proc. Natl. Acad. Sci. U. S. A. 97 (25), 13961–13966.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al., 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. J. Cogn. Neurosci. 12 (6), 1038–1055.

Pinker, S., 1994. The Language Instinct: How the Mind Creates Language. Penguin, London.

Poremba, A., Saunders, R.C., Crane, A.M., Cook, M., Sokoloff, L., Mishkin, M., 2003. Functional mapping of the primate auditory system. Science 299 (5606), 568–572.

Rauscheker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. 97 (22), 11800–11806.

Rivera-Gaxiola, M., Csibra, G., Johnson, M.H., Karmiloff-Smith, A., 2000. Electrophysiological correlates of cross-linguistic speech perception in native English speakers. Behav. Brain Res. 111 (1–2), 13–23.

Serniclaes, W., Sprenger-Charolles, L., Carre, R., Demonet, J.F., 2001. Perceptual discrimination of speech sounds in developmental dyslexia. J. Speech Lang. Hear. Res. 44 (2), 384–399.

Sharma, A., Dorman, M.F., 2000. Neurophysiologic correlates of cross-language phonetic perception. JASA 105 (5), 2697–2703.

Stevens, N.K., Blumstein, S.E., 1981. The search for invariant acoustic correlates of phonetic features. In: Eimas, P.D., Miller, J.L. (Eds.), Perspectives on the Study of Speech. Erlbaum, Hillsdale, NJ.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging. Translated by Mark Rayport. New York: Thieme Medical Publishers Inc. Stuttgart New York: George Thieme Verlag.

Tiitinen, H., May, P., Reinikainen, K., Näätänen, R., 1994. Attentive novelty detection is governed by pre-attentive sensory memory. Nature 372, 90–92.

Tremblay, K., Kraus, N., McGee, T., 1998. The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. NeuroReport 9, 35557–35560.

Vouloumanos, A., Kiehl, K.A., Werker, J.F., Liddle, P.F., 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J. Cogn. Neurosci. 13 (7), 994–1005.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al., 1999. Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36 (5), 638–642.

Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., Warburton, E.A., 2001. Separate neural subsystems within 'Wernicke's area'. Brain 124 (Pt 1), 83–95.

Zatorre, R.R., Binder, J.R., 2000. Functional and structural imaging of the human auditory system. In: Toga, A.W., Mazziotta, J.C. (Eds.), Brain Mapping: The Systems. Academic Press, San Diego, CA, pp. 365–402.

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., Meyer, E., Gjedde, A., Evans, A.C., 1996. PET studies of phonetic processing of speech: review, replication, and reanalysis. Cereb. Cortex 6 (1), 21–30.