Functional Segregation of Cortical Language Areas by Sentence Repetition

Ghislaine Dehaene-Lambertz,[1–5]* Stanislas Dehaene,[1,2,5] Jean-Luc Anton,[3] Aurelie Campagne,[3] Philippe Ciuciu,[4,5] Guillaume P. Dehaene,[1] Isabelle Denghien,[5] Antoinette Jobert,[1–5] Denis LeBihan,[4,5] Mariano Sigman,[1–5] Christophe Pallier,[1–5] and Jean-Baptiste Poline[4,5]

[1]*Unité INSERM 562, Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, Orsay, France*
[2]*Collège de France, Paris, France*
[3]*Centre IRMf, IFR 131, CHU Timone, Marseille, France*
[4]*UNAF, Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, Orsay, France*
[5]*IFR 49, Institut d'imagerie neurofonctionnelle, France*

◆ ━━━━━━━━━━━━━━━━━━ ◆

Abstract: The functional organization of the perisylvian language network was examined using a functional MRI (fMRI) adaptation paradigm with spoken sentences. In Experiment 1, a given sentence was presented every 14.4 s and repeated two, three, or four times in a row. The study of the temporal properties of the BOLD response revealed a temporal gradient along the dorsal–ventral and rostral–caudal directions: From Heschl's gyrus, where the fastest responses were recorded, responses became increasingly slower toward the posterior part of the superior temporal gyrus and toward the temporal poles and the left inferior frontal gyrus, where the slowest responses were observed. Repetition induced a decrease in amplitude and a speeding up of the BOLD response in the superior temporal sulcus (STS), while the most superior temporal regions were not affected. In Experiment 2, small blocks of six sentences were presented in which either the speaker voice or the linguistic content of the sentence, or both, were repeated. Data analyses revealed a clear asymmetry: While two clusters in the left superior temporal sulcus showed identical repetition suppression whether the sentences were produced by the same speaker or different speakers, the homologous right regions were sensitive to sentence repetition only when the speaker voice remained constant. Thus, hemispheric left regions encode linguistic content while homologous right regions encode more details about extralinguistic features like speaker voice. The results demonstrate the feasibility of using sentence-level adaptation to probe the functional organization of cortical language areas. *Hum Brain Mapp 27:360–371, 2006.* © **2006 Wiley-Liss, Inc.**

Key words: MRI; brain; voice; speech; habituation; asymmetry; auditory cortex

◆ ━━━━━━━━━━━━━━━━━━ ◆

INTRODUCTION

Linguistic tasks, from syllable discrimination to story listening, activate a large network of perisylvian areas. Contrary to neuropsychological studies, in which specific deficits, such as articulatory deficits [Dronkers, 1996], phonological deficits [Caplan et al., 1995; Caramazza et al., 2000], and semantic deficits [Hart et al., 1985], have begun to dissect the brain circuitry involved in language, brain imaging studies have been relatively less successful in the segregation of functionally distinct regions within this perisylvian network. For example, activations are detected in the left inferior frontal gyrus and insula in syllable discrimination tasks [Dehaene-Lambertz et al., 2005; Zatorre et al., 1992], metaphonological tasks [Burton et al., 2000], syntactical processing [Hashimoto and Sakai, 2002; Kaan and Swaab, 2002],

and sentence listening [Dehaene et al., 1997; Pallier et al., 2003]. Although careful subtractions between adequate conditions have identified clusters within this network, their precise involvement in the studied task is still under question after more than a decade of brain imaging research. In this article we tackle this question from another angle, combining timing information to identify clusters with different dynamic and functional properties within the cortical language network and the priming method.

The adaptation or priming method is based on the observation that stimulus repetition induces a decrease in brain activity (repetition suppression), which can be measured with event-related potentials (ERPs) [Dehaene-Lambertz and Dehaene, 1994] and functional MRI (fMRI) [Grill-Spector et al., 1998; Naccache and Dehaene, 2001]. More importantly, the same decrease in activity can be observed when the repeated stimuli are not exactly similar but only share a common property that is extracted at the probed cortical location. This method is thus sensitive to the code used in a particular brain region. By varying the property that is repeated, it becomes possible to target a specific representation and to uncover its cerebral bases. Priming designs have been successfully used to characterize the processing steps in visual object perception [Grill-Spector and Malach, 2001], reading [Dehaene et al., 2004], or number representation [Naccache and Dehaene, 2001; Piazza et al., 2004]. In this approach, the first step consists in isolating the network involved in the object representation by studying which brain regions display repetition suppression when the same object is repeated. Then in successive experiments shared properties between the repeated stimuli are varied in order to identify which coding variations are relevant and which are not relevant to obtain a repetition suppression effect in a given brain region.

Here we applied this reasoning to speech perception. In the first experiment, we aimed to isolate which brain regions are sensitive to the mere effect of repetition of the same sentence. In the second experiment, analyzed in the Functional Imaging Analysis Contest (FIAC), we moved forward to characterize the functional properties of these perisylvian regions by repeating only one property of the sentences, either the sentence content or the carrier of this content: the voice of the speaker.

EXPERIMENT 1

In this experiment, sentences were repeated two to four times in a row in a slow-event design. Our goal was to examine whether the linguistic network can be decomposed into spatially organized functional circuits based on the timing of their BOLD response and on the dynamics of their adaptation. Previous priming experiments used brief presentations in the visual domain (faces, objects, words, and numbers), but repetition suppression has been described in auditory neurons [Ulanovsky et al., 2003] as well as in visual neurons [Miller et al., 1991]. Furthermore, repetition of syllables and pseudowords induces a decrease in the amplitude of evoked responses or MRI activations [Belin and Zatorre,

2003; Cohen et al., 2004; Dehaene-Lambertz and Gliga, 2004], suggesting that repetition suppression may be a general property of the brain, not limited to visual perception. Nevertheless, sentences typically last a few seconds and multiple processes at different integration levels (acoustic, phonetic, lexical, syntactic, local semantic, contextual semantic) are required to perceive their similarity. The observation of repetition suppression effects at a long lag (14.4 s) would suggest that the priming method can also target such long time range representations.

Methods

Participants

Ten right-handed young French adults (5 women and 5 men, ages 21–35 years), with no history of oral or written language impairment, neurological or psychiatric disease, nor hearing deficits were tested. All participants gave their written informed consent and the study was approved by the local ethics committee.

Stimuli

A female French speaker was recorded reading a well-known children's story, "The Three Little Pigs." People know the characters and the general scenario of this story. However, sentences were rewritten and thus were original. The advantage of a children's story is to elicit attention from the listener because of frequent changes of register (dialogue, exclamation, description, etc.). The speaker used a theatrical intonation to reinforce this aspect. Sixty sentences with a mean duration of 2200 ms (1580–2863 ms) were extracted from the original story. The overlearned and expected sentences of the story were not presented. More details on the stimuli can be found in Dehaene-Lambertz and Houston [1998].

Experimental procedure

A randomly chosen sentence was presented every 14.4 seconds in a slow-event-related design. Each sentence was then repeated two, three, or four times in a row still with a 14.4-second intersentence interval. Each participant listened to a total of 180 sentences, 20 different sentences in each of the three repetition conditions. These repetition conditions were randomly ordered. The experiment was divided in five runs, each comprising 36 sentences and lasting 8 minutes 30 seconds. The participants passively listened to the sentences. To keep them minimally attentive during the experiment, they were instructed that they would have to recognize the story from which the sentences were extracted and to answer questions about the sentences at the end of the experiment. All participants recognized the story.

Image acquisition and analysis

The experiment was performed on a 3T 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 (26 contiguous axial slices, thickness 4.5 mm, TR = 2.4 s, TE = 40 ms, flip angle = 90°, field-of-view = 192 × 256 mm, 64 × 64 pixels). A high-resolution ($1 \times 1 \times 1.2$ mm^3), T1-weighted 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/). For each participant, after image reconstruction of the functional runs the first four volumes of each run were discarded to eliminate nonequilibrium effects of magnetization. The remaining volumes were corrected for the temporal delays between slices and realigned to the first volume. The mean realigned image was used to check the correct alignment of the functional images with the structural image. The T1-weighted structural MRI scan was spatially normalized to the Montreal Neurological Institute (MNI) template and the normalization parameters were applied to the functional images. These images were then resampled every 4 mm using bilinear interpolation and spatially smoothed with an isotropic Gaussian filter (kernel = 5 mm). The time series in each voxel was highpass-filtered (cutoff = 60 s), smoothed with a 4-s Gaussian kernel, and globally normalized with proportional scaling.

A linear model was generated by entering, for each run, four distinct variables corresponding to the four repetition positions of a sentence: first, second, third, and fourth presentations. Due to the paradigm construction, there were 60 events for the first and second presentation, 40 for the third presentation, and only 20 for the fourth presentation. The variables convolved by the standard SPM hemodynamic response function (HRF) and their temporal derivatives were included in the model. For random effect group analyses, the contrast images from individual participant's analysis were smoothed with an 8-mm Gaussian kernel and submitted to one-sample t-tests. We first examined the activations separately for the first to fourth presentations of each sentence. Second, we tested whether activations for a particular sentence repetition significantly decreased relative to the preceding sentence presentation. Finally, we tested for a significant linear decrease across repetition from the first presentation of the sentence to the fourth.

In order to better estimate the periodicity and phase of the event-related BOLD response, a distinct model was also designed by convolving the same four variables with a single cycle of a sine and a cosine waveform at a frequency of 14.4 s. The ratio of the regression weights of the sines and cosines was then transformed with the inverse tangent function to yield a BOLD response phase between 0 and 2π. This phase was multiplied by the stimulation period (14.4 s) and divided by 2π to yield a phase lag expressed in seconds. BOLD response amplitude was computed as the square root of the sum of the squares of the sine and cosine coefficients. To estimate the presence of a significant response at the stimulation period, an analysis across sessions was performed using the Rayleigh circular statistic, which contrasts the observed distribution of phases to the null hypothesis of a uniform distribution across the phase circle. To assess the presence of repetition effects, the images of amplitude and phase computed for each participant and for each repetition position were entered in separate random-effect group analyses and submitted to one-sample t tests.

If not reported otherwise, all the reported effects passed a voxelwise threshold of $P < 0.001$ uncorrected for multiple comparisons and a $P < 0.05$ threshold on the extent of clusters.

Results

Activations to the first presentation of a sentence involved both superior temporal gyri and sulci, the left inferior frontal region and insula, both thalami, and the right caudate nucleus, whereas for the fourth presentation, activations were limited to the middle part of the superior temporal gyrus, anteriorly and posteriorly to Heschl's gyri. As illustrated in Figure 1, a decrease in amplitude was already observed for the second presentation of the sentence and was significant in both temporal regions and in the left inferior frontal gyrus and insula (Table I). Although additional decreases with successive repetition were small, a significant linear effect was present mainly over the left superior temporal region (252 voxels, Fig. 1), while on the right side the linear effect of repetition was limited to 50 voxels (Table I). Some regions, such as Heschl's gyrus and its vicinity, presented no effect of repetition (Fig. 1).

We analyzed also the phase and periodicity of the BOLD response: Two effects are apparent in the hemodynamic response curves shown in Figure 1. First, the latency of the BOLD response seems to vary across regions, and second, it seems to decrease with sentence repetition. To rigorously quantify both effects, we fitted the whole-brain data with a set of sine and cosine functions in order to extract the phase and amplitude of the BOLD response (see Methods, above). We then submitted both parameters to a between-participants random effect analysis. The amplitude data essentially replicated the above SPM analysis, showing both a strong activation to the first sentence, in bilateral temporal and left inferior frontal cortex, as well as an adaptation effect particularly evident from the first to the second repetition. The phase analysis, however, yielded novel evidence in favor of a hierarchical temporal organization of these areas. In the image of the phase of the response to the first sentence, we observed, in all participants, a temporal progression with the earliest responses observed in Heschl's gyrus, followed by successively slower responses in posterior and middle STS, temporal pole, and inferior frontal gyrus. The phase averaged over subjects to the first and to the second sentence (converted in seconds) is shown in Figure 2. As illustrated in this figure, in several regions the phase showed a slight acceleration with sentence repetition. A random effect analysis was used to compare the phases to the first sentence and to the mean of the phases of repetitions 2–4. Search was confined to the network activated by the presentation of the first sentence (Fig. 1), and to increase sensitivity, the voxel threshold was decreased to $P < 0.01$ (still with corrected

Figure 1.
Repetition suppression related to sentence repetition. **A:** In Experiment 1: sagittal slice at x = −49 mm (standard Talairach coordinates) displaying the regions in which repetition induces a linear decrease in amplitude (blue-yellow scale), superimposed on the activations to the first sentence (yellow-red scale). Graphs show the adaptation of the mean BOLD response with sentence repetition at four locations that illustrate the different patterns of responses to sentence repetition. Note also the different shapes of the BOLD responses, with, for example, an earlier peak in the middle STS than in the posterior STS. Coordinates are given in standard Talairach coordinates. **B:** In Experiment 2: sagittal slice at x = −55 mm (Talairach coordinates) displaying the regions significantly more activated when both parameters, sentences and speakers, varied than when both were repeated (blue-yellow scale) superimposed on the activations to the first sentence (yellow-red scale).
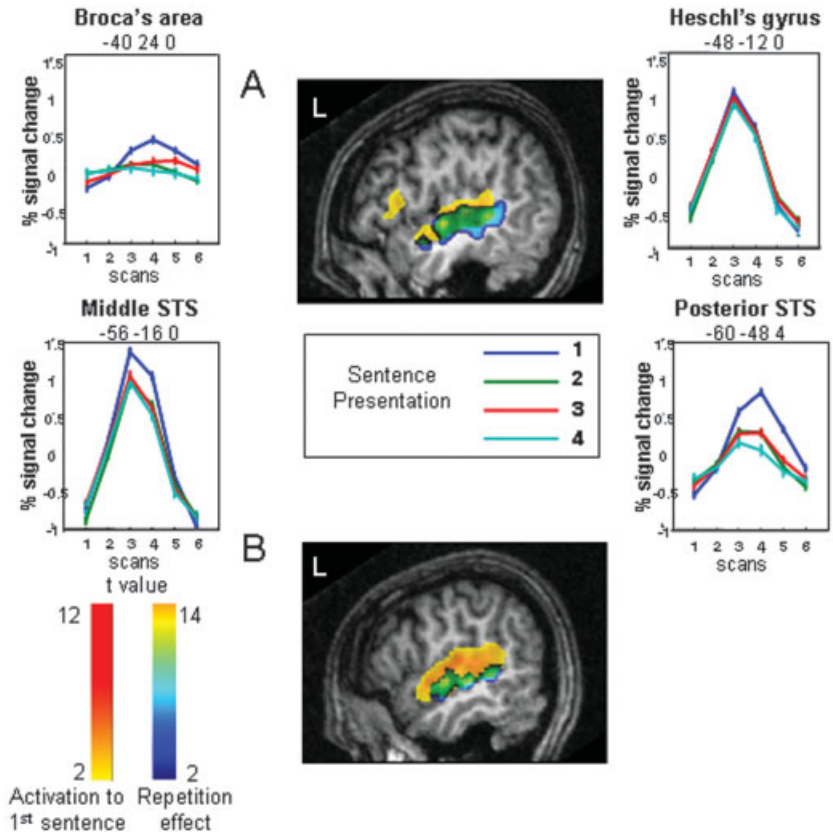
cluster-level $P < 0.05$). This analysis revealed a significant acceleration with repetition only in the right STS (Talairach coordinates: 60, −12, 0; Z = 3.75 and 56, −24, 0; Z = 2.96), and a marginally significant effect in the left STS (−56, −8, 0; Z = 3.43 and −56, −20, 0; Z = 3.24).

Discussion

Using a slow event-related paradigm, we observed a hierarchical temporal organization along the superior temporal regions with the earliest response in the primary auditory cortices and the slowest in associative regions with a caudal–rostral and a dorsal–ventral gradient along the superior temporal sulcus (Figs. 1, 2). Repetition of sentences affected both amplitude and phase of the HRF response. This adaptation effect was particularly evident between the first and second sentence, although it continues with the following repetitions. The pattern of adaptation was different across regions (Fig. 1) with a set of regions demonstrating the same response each time a sentence was presented (e.g., Heschl's gyrus), regions showing a strong decrease between the first and second presentation and more subtle decrease with the following repetitions (e.g., Broca's area) and regions that show a close to linear adaptation with repetition (e.g., superior temporal sulci).

The phase delay between the fastest region (Heschl's gyrus) and the slowest (temporal poles and inferior frontal regions) varies from 2 to 8 s. This delay cannot be related to

an acquisition delay between the different slices. First, we applied a temporal correction for slice acquisition times to the original data. Second, acquisition was from bottom to top, whereas the slowest phases are observed in the more basal slices. Third, the observed delays are much larger that the slice acquisition lags and, in several cases, multiple delays are seen in the same slice.

Several studies have already reported delays in the time course of activations in frontal areas relative to posterior areas [Schacter et al., 1997; Thierry et al., 1999]. Note that here we are measuring phase, a measure sensitive not only to the rise of the hemodynamic response, and thus to the peak latency as measured in previous experiments [Henson et al., 2002; Schacter et al., 1997; Thierry et al., 1999, 2003], but also to its plateau (possibly including neural responses posterior to sentence offset). This methodological difference may explain the large interval of phase lags observed (2–8 s). Although hemodynamic differences between cortical areas may contribute to the observed differences, the observed delays seem too large to be due solely to factors other than differences in cognitive processing. Indeed, several previous experiments also showed that BOLD response delays could be affected by cognitive factors. By comparing brain activity evoked by decision tasks bearing on the first or second presentation of a pair of nouns, Thierry et al. [2003] demonstrated that they could further delay the peak of the BOLD response in frontal areas. Conversely, Henson et al. [2002]

TABLE I. fMRI activations in Experiment 1

| Area	No. voxels in cluster	Cluster-level P value (corrected)	Z value at local maximum	Talairach coordinates: x, y, z
1. First sentence				
Right STG and STS	332	<0.001	5.51	44, −28, 0
			4.57	60, −24, 4
			4.04	40, −48, 16
Left STG and STS	253	<0.001	5.15	−52, −32, 4
			3.88	−48, 0, −12
			3.50	−60, −56, 12
Left inferior frontal and insula	61	<0.001	4.02	−48, 16, 8
Thalami	140	<0.001	3.93	−16, −20, 16
			3.87	8, −12, 12
Right caudate nucleus	37	0.002	3.78	24, 24, 4
2. Fourth sentence				
Right STG	147	<0.001	4.23	48, −36, 0
			3.54	36, −20, −8
Left STG	78	<0.001	4.38	−60, −32, 4
			4.28	−52, −16, 0
3. First > second presentation				
Right STG and STS	252	<0.001	5.35	60, −36, 4
			5.12	48, −48, 8
			4.11	52, −12, −4
Left STG and MTG	147	<0.001	4.91	−52, −20, 0
			4.65	−60, −44, 4
Left insula	32	.02	3.67	−32, 16, 8
4. Regions displaying linear habituation				
Right STG and STS	50	0.005	3.92	44, −32, 4
			5.12	48, −16, 0
Left STG and STS	223	<0.001	5.27	−56, −16, 0
			4.79	−60, −48, 8

STG: superior temporal gyrus; STS: superior temporal sulcus; MTG: middle temporal gyrus.

studied repetition effects in face processing using a design comparable to the present one, and observed that the second presentation of a face elicited an earlier response in the right fusiform area. Similarly, in the present experiment sentence repetition accelerated the phase of the BOLD response, which would not be expected if the delays were due solely to hemodynamics.

While previous studies observed solely a difference in BOLD response delay between frontal regions and temporal regions [Schacter et al., 1997; Thierry et al., 2003], we demonstrate here a gradual organization of phases along the superior temporal regions with a slowing down extending in both anterior and posterior directions from Heschl's gyrus. Along with this anterior–posterior gradient, there is also a dorsal–ventral gradient with fastest phase along the sylvian fissure than along the superior temporal sulcus. These gradients are compatible with the known organization of auditory connections leading from the superior temporal region towards the temporal pole and the ventrolateral prefrontal cortex [Petrides and Pandya, 2002], through distinct ventral and dorsal pathways [Romanski et al., 1999]. The observed fMRI delays in excess of 1 or 2 s are unlikely to reflect the transmission of neural information from one region to the next, which occurs on the scale of a few tens of milliseconds. This gradient might rather be the result of different cognitive operations that integrate larger and pos-

sibly more abstract speech units that may require longer processing time, and/or with a more sustained activity.

This hypothesis of hierarchical integration of larger units along the superior temporal regions is confirmed by the repetition suppression effect. Regions coding segmental properties (acoustical and phonological) were expected to show no effect of repetition contrary to those processing supra-segmental properties (prosodic, semantic, and syntactic properties). Indeed, a similar anterior–posterior, dorsal–ventral gradient was observed for repetition effects, with both superior temporal sulci strongly affected by repetition. Roughly three types of adaptation patterns can be described: The fastest regions, Heschl's gyrus and its vicinity in the superior temporal gyrus, are not affected by repetition, either because they are unable to code an entire sentence, or to maintain this code during the time lag of 14.4 s separating two repetitions. The fast response of this region, its insensitivity to sentence repetition, confirms the role of segmental coding proposed for this auditory region with a time unit of a few tens or hundred milliseconds [Boemio et al., 2005]. At the opposite, Broca's area, one of the areas showing the slowest phase, is significantly activated only for the first presentation, displaying a sharp adaptation effect. Although Broca's region has been classically associated with syntactic processing, its role in syntax is now reassessed and discussed in a more general framework of working memory
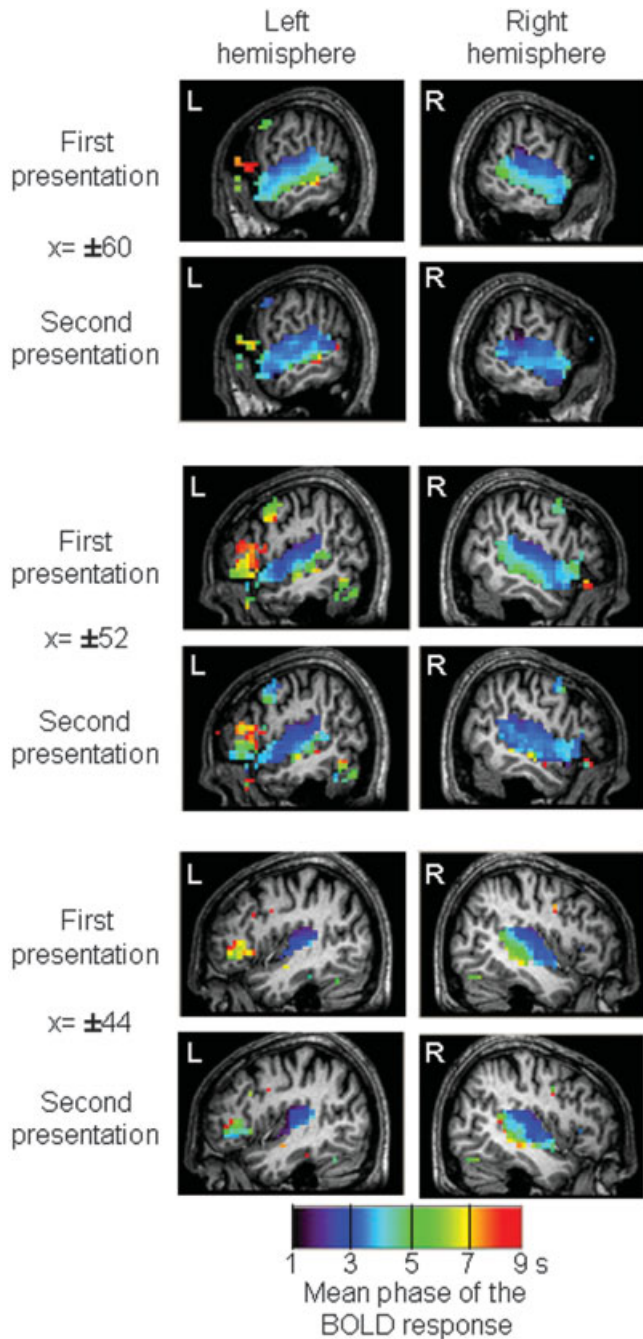
Figure 2.
Temporal organization of cortical responses to sentences for the first and second presentation of the same sentence. Colors encode the circular mean of the phase of the BOLD response, expressed in seconds relative to sentence onset. Fastest responses, in purple, are visible in Heschl's gyrus, while the slowest responses, in yellow and red, are encountered in Broca's area. A delay in phase is visible along the superior temporal region with a dorsal–ventral and anterior–posterior gradient. The second presentation of a sentence speeds up the phase in all these regions.

load [Kaan and Swaab, 2002]. In our experiment, the syntax was highly varied across sentences but the sentences were short, with no particular syntactical ambiguities or difficulties. A significant activation present in this region only for the first presentation suggests two different hypotheses. First, once the syntactic tree is computed for the first presentation, it might be easily reapplied to successive presentations with no renewed computations. Second, this region might be incidental in sentence perception and might be recruited only when participants have time to rehearse the sentence in order to integrate it in the global context of the story. This single experiment cannot separate these hypotheses.

Finally, temporal regions with a median phase response showed a clear effect of repetition suppression but remained significantly activated even when the sentence was presented for the fourth time. The decrease of the amplitude of the BOLD response and the speeding-up of the phase (Figs. 1, 2) along the superior temporal region might reflect either a general decrease of attention and/or an improved contextual integration of the sentence elements in the local context of the sentence and in the global context of the story.

To summarize, repetition induced a decrease in amplitude and a speeding up of the phase of the BOLD response with a gradient roughly similar to what is obtained by measuring the phase of the BOLD response. In order to deepen our understanding of the role of these regions, our second step involved manipulating which parameter was repeated. Among the regions that show repetition suppression in Experiment 1, can we separate those sensitive to the linguistic content and those sensitive to the carrier of this content, i.e., the speaker's voice?

EXPERIMENT 2

A spoken utterance conveys not only linguistic information but also information about the speaker identity and their emotional state. Each individual possesses voice characteristics due to the configuration of his/her vocal tract, his/her pronunciation, his/her dialectal accent making him/her recognizable even when the listener does not previously know the speaker. The capacity to identify speaker voice is present very early on. Infants [Mehler et al., 1978], even fetuses [Kisilevsky et al., 2003], are able to recognize their mother and to discriminate stranger's voices [Dehaene-Lambertz, 2000].

Classical theories of speech perception and word recognition postulate the existence of an early processing stage of "speaker normalization" that deletes voice-specific features from the acoustic signal. Yet some researchers believe that such a processing stage is not necessary and have proposed that word recognition proceeds by comparing the acoustic input with multiple exemplars (or templates) stored in lexical memory [Goldinger, 1998]. Features of the speaker voice are assumed to remain encoded in those exemplars. Evidence in favor of this claim comes from word identification experiments using repetition priming that showed that a given word is better identified when it is repeated with

similar acoustic features (e.g., when spoken by the same speaker) than when it is repeated with different acoustic details [Church and Schacter, 1994; Goldinger, 1996].

Brain imaging provides the opportunity to disentangle brain regions that are sensitive to the repetition of abstract linguistic content, and regions that are sensitive to the repetition of more detailed acoustic features. The goal of our second experiment was thus to separate surface and abstract representations of sentences by using a repetition paradigm. We manipulated independently the repetition of two parameters: speaker and sentence content. Either the same sentence was repeated but said by different speakers (SSt-DSp) or the speaker was constant but produced different sentences (DSt-SSp). These conditions were compared with two other conditions in which the same sentence produced by the same speaker was repeated several times (SSt-SSp), and different sentences were produced by different speakers (DSt-DSp). We expected a decrease of activity in regions that code for linguistic information when the same sentence was repeated. If there is normalization, the same decrease should be present when the speaker is held constant or when it is varied. Similarly, regions that code for speaker identity should show a decrease when the speaker is held constant, even when the linguistic content is varied. In order to avoid an exceedingly long experimental duration, we used short blocks of six sentences (block-design) that were either all similar or all different for the parameter of interest.

Methods

Participants

Sixteen right-handed young French adults with no history of oral or written impairment, neurological or psychiatric disease, or hearing deficits were participants. They gave their written informed consent, and the study was approved by ethics committee. Acquisition was stopped in Participant 5, leaving 15 participants complete the data set (8 women and 7 men; mean age, 27 years; range, 21–35).

Stimuli

A female French speaker was recorded reading a well-known children's story, "The Three Little Pigs," with a theatrical intonation. Sentences were rewritten and thus were original. We extracted 64 sentences from the same story used in Experiment 1. Each sentence was secondarily recorded by 9 other French speakers with the instruction to not only repeat the sentence but to follow as much as possible the intonation and rhythm of the first speaker. We obtained 640 sentences; i.e., 64 sentences produced by 5 men and 5 women with a mean duration of 2277 ms (1515–2793 ms).

Experimental procedure

The experiment was divided into four runs, two runs with a block design and two with an event-related design. The order of the four runs was counterbalanced across participants. In both designs, the instructions given to the participants were to

lie still in the scanner with eyes closed and to attentively listen to the sentences, because they would be asked whether sentences presented after scanning were presented during the experiment. Only the block design is analyzed below, whereas the event-related design is analyzed in other papers in this special issue (see Poline et al., Table I).

Block design. The sentences were presented in small blocks of 6 sentences, one every 3333 ms (thus the silence between sentences was variable because of the variation in sentence length). There were four types of blocks: Same Sentence-Same Speaker (SSt-SSp), the same sentence said by the same speaker was repeated 6 times; Same Sentence-Different Speakers (SSt-DSp), the same sentence was repeated by 6 different speakers (3 men and 3 women); Different Sentences-Same Speaker (DSt-SSp), the same speaker produced 6 different sentences; Different Sentences-Different Speakers (DSt-DSp), 6 different speakers (3 men and 3 women) produced 6 different sentences. For each block, the speakers were randomly selected among the 10 original speakers with the constraint that male and female voices were equally distributed within each block and across blocks. For all conditions, sentences were randomly selected on line with the constraint that a sentence could not be repeated in the first 8 blocks. The selection of the sentences for the condition (SSt-DSp) was limited to a subset of 96 sentences (16 sentences × 6 speakers) in which prosody was the most homogeneous among speakers. The order of speakers was randomly selected for each block in which different speakers were present. Each block lasted 20 seconds and was followed by a silence of 9 seconds. The order of the conditions was randomly selected for each participant with the constraint that no more than two blocks of the same condition were presented in a row. Across the two runs used in the block design, a total of 8 blocks were obtained in each condition presented (4 blocks per run and per condition).

Event-related design. One sentence was presented every 3333 ms. The same conditions as above were presented but were defined in this design by the transition between two sentences. Four types of transitions were thus possible: The sentence was similar to the previous one (SSt-SSp), or there was a change of speaker (SSt-DSp), or a change of sentence (DSt-SSp), or a change of sentence and speaker (DSt-DSp). Note that the change of speaker could be either within the same sex, or across sexes. Participants performed two runs of 141 sentences each giving 72 transitions in each condition.

Image acquisition

The experiment was performed on a 3-T whole body system (Bruker, Germany), equipped with a quadrature birdcage radio frequency (RF) coil. Functional images comprising 30 axial slices (thickness = 4 mm) covering most of the brain were obtained with a T2-weighted gradient echo, EPI sequence (interleaved acquisition with the following parameters: TR, 2.5 s; TE, 35 ms; flip angle, 80°; field-of-view, 192 × 192 mm; 64 × 64 pixels). A high-resolution (1 × 0.9 × 1.4 mm), T1-weighted, anatomical image using a 3-D gradi-

TABLE II. fMRI activations in Experiment 2

Area	No. voxels in cluster	Cluster-level P value (corrected)	Z value at local maximum	Talairach coordinates: x, y, z
1a. Main effect of sentence repetition				
Left middle STS	61	0.016	4.21	−63, −15, 0
Left posterior MTG	25	0.246	3.46	−57, −42, 0
1b. Asymmetries in the main effect of sentence repetition				
Left posterior MTG	89	<0.001	3.96	−54, −42, 0
Left middle STS	15	0.342	3.57	−60, 15, 6
2. Main effect of speaker repetition				
No significant regions				
3. Interactions speaker × sentence repetitions				
No significant regions				
4a. Effect of sentence repetition restricted to same speaker condition				
Left middle STS	49	0.055	4.54	−63, −12, −3
Left posterior MTG	62	0.024	4.07	−60, −42, 0
4b. Asymmetries in the effect of sentence repetition restricted to same speaker condition				
Left posterior MTG and STS	70	0.001	3.99	−57, −42, 0
5a. Effect of sentence repetition restricted to different speaker condition				
Left middle STS	52	0.023	3.84	−63, −15, −3
5b. Asymmetries in the effect of sentence repetition restricted to different speaker condition				
Left middle STS	41	0.001	4.62	−57, −12, 6
Left posterior MTG	46	0.006	3.90	−60, −39, 0

STG: superior temporal gyrus; STS: superior temporal sulcus, MTG: middle temporal gyrus.

ent-echo inversion-recovery sequence was also acquired for each participant.

Image analysis

fMRI data analysis was performed using Statistical Parametric Mapping (SPM2, http://www.fil.ion.ucl.ac.uk/spm/). Preprocessing was the same as Experiment 1 except that the images were resampled every 3 mm and that the time series in each voxel was highpass-filtered with a cutoff of 128 s.

Block design analysis. For each participant a linear model was generated by entering, for each run, five distinct variables corresponding to the first sentence pooled across all conditions and to the second to sixth sentences, separately for the four conditions. The logic was that the conditions differed only after the second sentence and that the response to the first sentence might differ because it follows a long period of silence. The variables were convolved by the standard SPM hemodynamic response function (HRF). For random effect group analyses, the individual contrast images from individual participants were smoothed with an 8-mm Gaussian kernel and submitted to one-sample t-tests. The design being a 2 × 2 factorial design, we computed the main effects of sentence and of speaker and their interactions. From a cognitive point of view, we were interested in the brain regions that were sensitive to sentence repetition even when acoustic variability due to a change of speaker was present, and to the brain regions sensitive to speaker repetition when sentences were constant or varied. Thus, we also computed the corresponding four contrasts.

We tested also for significant left–right asymmetries in these analyses: For each participant the transformation matrix of the anatomical image toward its flipped image by a nonlinear normalization was computed and then applied to the smoothed contrast images of the previous analyses to obtain their flipped images. Individual asymmetry images were obtained by subtracting the original contrast image from its flipped version. Those images were then 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.

If not reported otherwise, all the reported effects passed a voxelwise threshold of $P < 0.001$ uncorrected for multiple comparisons and a $P < 0.05$ threshold on cluster extent.

Results

The first sentence of all blocks activated bilateral superior temporal regions and the left posterior part of the brainstem (colliculi). A significant asymmetry favoring the left side was observed in the posterior part of the superior temporal sulcus extending dorsally and medially over the planum temporale (272 voxels, Z = 4.75 at $x = −39$, $y = −36$, $z = 12$, and Z = 4.7 at $x = −57$, $y = −39$, $z = 6$).

A main effect of sentence repetition was observed in the middle part of the left temporal sulcus. This region became habituated even when speakers varied (DStDSp-SStDSp), demonstrating a normalization property. A significant asymmetry favoring the left side was present for the latter comparison in this cluster and in a second more posterior cluster located in the middle temporal gyrus (Table II).
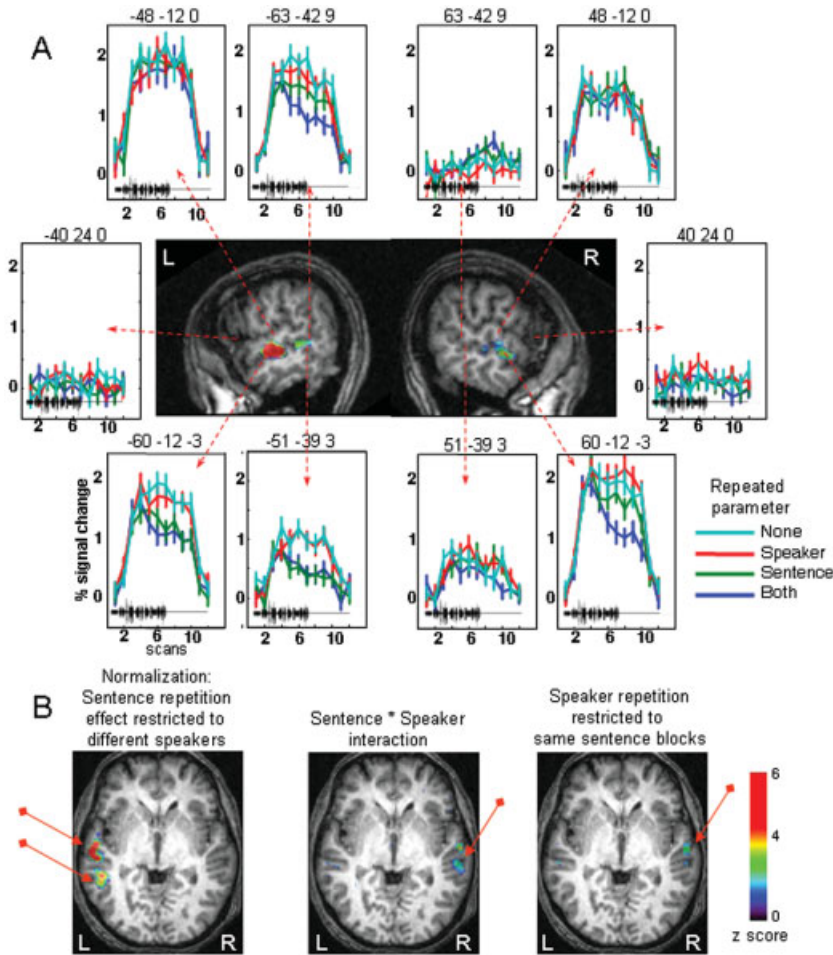
Figure 3.

Repetition suppression related to speaker and sentence repetition in Experiment 2. **A:** Graphs, surrounding sagittal slices at $x = \pm 60$ mm (standard Talairach coordinates), show the adaptation of the mean BOLD response in the different conditions. The black tracing on each plot represents the duration of the speech stimuli (~19 s). Several patterns are seen. The superior temporal regions (e.g., -48 mm, -12 mm, 0 mm) are activated and do not adapt, while inferior frontal regions (-40 mm, 24 mm, 0 mm) are inactive. The left STS show adaptation to sentence repetition, even when speakers varied (e.g., -60 mm, -12 mm, -3 mm). Right homologous regions either did not adapt (e.g., 51 mm, -39 mm, 3 mm) or adapted only when the speaker was maintained constant (60 mm, -12 mm, -3 mm). **B:** Axial slices at $z = 0$ mm (Talairach coordinates), displaying the clusters activated in the different comparisons.

This classical SPM analysis did not reveal any effect of speaker repetition nor significant interaction between speaker and sentence repetition factors. There were also no significant asymmetries for these contrasts. However, more sensitive analyses based on permutations and mixed-effect model, as presented by Meriaux et al. [2006], found a significant main effect of speaker repetition in the left superior temporal sulcus ($x = -63, y = -42, z = -9$) and a significant interaction between Sentence and Speaker repetition in the right middle temporal gyrus ($x = 60, y = -12, z = -3$) (Fig. 3; see tables in Meriaux et al. [2006]). The latter region significantly habituated when the speaker was kept constant relative to blocks of varied speakers; however, it did so only when the sentence was maintained constant (SStDSp-SStSSp). In the latter comparison, a second cluster was also significant in the left superior temporal gyrus ($x = -60, y = -30, z = 6$).

Discussion

In this second experiment where a change of sentence was contrasted with a change of speaker, we observed a reliable effect of adaptation to the linguistic content in the left superior temporal sulcus that was independent of the speaker, while the converse adaptation to speaker identity was not observed. Adaptation to sentences was significantly asymmetric, present only in the middle and posterior left temporal region, while the contralateral right regions were unable to normalize across speakers. The right temporal region habituated to sentence repetition only when the speaker was maintained constant.

Speech normalization

Speech is produced by different vocal tracts inducing variations in the acoustical cues that support linguistic content. Normalization of speech perception across many different speakers has long been recognized as one of the crucial difficulties that the brain has to resolve. Liberman [1985] proposed that normalization across speakers is realized through the recognition of the motor pattern or gesture that underlies the overt phoneme realization. The recent discovery of mirror neurons in a possible equivalent of Broca's area in the macaque has given biological support to the notion of a common representation between auditory or visual percepts and motor patterns [Kohler et al., 2002]. However, we found no evidence of Broca's involvement in the normalization process engaged here (Fig. 3). Because the

same sentences were used in Experiments 1 and 2, the activation that we found in Broca's area in Experiment 1 might be related to the participants' rehearsal of the sentence, which is possible during the pauses of the slow-event paradigm. The incidental involvement of the left inferior frontal region in sentence comprehension is further underscored by a transcranial magnetic stimulation (TMS) study showing that magnetic stimulation over that region had no effect on a participant's sentence perception, contrary to the same stimulation over the posterior temporal area that facilitates native language processing [Andoh et al., 2006]. These results are in contradiction to those reported by Sakai et al. [2002] and Hashimoto and Sakai [2002] underscoring the involvement of the left inferior frontal gyrus in sentence comprehension. However, as the authors pointed out, the involvement of this region in speech comprehension was demonstrated only when the task explicitly requires the use of syntactic rules or when participants have to recover words in degraded speech [Davis and Johnsrude, 2003]. Broca's area might thus be involved in speech listening when there is a conscious effort in sentence processing to integrate the different elements of the sentence.

The only regions displaying a normalization property were in the superior temporal sulcus, a unimodal auditory region [Poremba et al., 2003] that reacts more to speech than to other auditory stimuli (see Binder et al. [2000] for a meta-analysis of speech vs. nonspeech studies). In particular, using sinewave speech we observed that an area encompassing the two maxima observed here was more activated when stimuli were perceived as speech than when the same stimuli were perceived as whistles [Dehaene-Lambertz et al., 2005]. The same clusters are also activated by different types of degraded speech but only when they are intelligible [Davis and Johnsrude, 2003; Narain et al., 2003]. These results underscore that this part of the STS is not sensitive to the surface form of the auditory stimuli, but rather to the linguistic representations elicited by them. In an experiment of word repetition contrasting oral and visual modalities, Cohen et al. [2004] proposed that an equivalent of the visual word form area might exist in the superior temporal sulcus. This auditory word form area (AWFA), which showed repetition suppression within the auditory modality but not across different modalities, would be tuned to recognize words in the auditory environment irrespective of irrelevant features, such as speaker identity, pitch, speech rate, etc. It is noteworthy that the tentative coordinates that they propose for the AWFA (−60, −8, −4) fall close to the main peak of speaker independent adaptation observed in the present study (−63, –15, –3). We also observed a second spot presenting these characteristics, posterior to the auditory cortex. Based on this experiment we cannot conclude whether the code unit in these regions is limited to words or whether it can integrate an entire sentence. However, one might speculate, following the classical distinction between dorsal and ventral pathways [Hickok and Poeppel, 2000; Scott and Johnsrude, 2003], that these two clusters might be associated with different coding schemes: the anterior region would map the word form onto lexical representations, while the posterior region, through the dorsal pathway, would maintain the phonological word form in order to interface with the working memory and the motor systems. In a similar priming study, but using written sentences, Noppeney and Price [2004] reported syntactic adaptation in the left temporal pole (Talairach coordinates: −42, 3, −27). We could have missed this region because of magnetic susceptibility artifacts in the temporal pole. However, the major difference between both studies, aside from the visual vs. auditory modality, is that the depth of syntactic computations might be different in both tasks. Similarity between sentences could be detected at several levels here. Our sentences were also simple to understand and never comprised ambiguous syntactic structures that might have pushed the participants in Noppeney and Price's study to rely more on syntactic processing in order to understand the sentences even in nonambiguous cases (see Hahne and Friederici [1999] for the effect of the proportion of syntactically ambiguous sentences).

Hemispheric asymmetry

The functional asymmetry between the two superior temporal sulci is striking. Repetition suppression is observed in left and right homologous clusters in the middle superior temporal sulcus. However, on the right side this effect is limited to the condition where both the sentence and the speaker remain constant. Contrary to the left cluster, the right cluster does not possess normalization capacities. These complementary codes, one abstract on the left side, the other exemplar-dependent on the right side, might explain performances in behavioral priming experiments. Lexical decision, which is primed by previous presentations of the word, even produced by different voices [Luce and Lyons, 1998], would be based on the left abstract code while explicit word recognition, which is affected by voice differences [Church and Schacter, 1994], might be informed by both regions. This left–right difference is reminiscent of another similar asymmetry described in the fusiform area for visual stimuli. Whereas the left fusiform area displays repetition suppression when a different view of the object is repeated, the homologous right region shows repetition suppression only when the same view is repeated [Vuilleumier et al., 2002]. The authors argued that this difference was independent of lexico-semantic factors but related to view-invariant properties of the visual code in the left hemisphere. In both visual and auditory perception the left hemisphere appears to compute an abstract categorical representation, stripped of sensory details, while the right homologous region maintains more surface details.

Adaptation to the speaker

While we observed regions that habituated to sentence, we did not find regions that habituated to speaker repetition. In our experiment, although the participants did not know the speakers, we could have expected a difference because blocks with different speakers comprise male and

female voices, contrary to blocks with the same speaker. The right anterior STS has been identified as an important region involved in identifying speakers [Belin et al., 2004; von Kriegstein and Giraud, 2004]. In an adaptation paradigm, Belin and Zatorre [2003] compared blocks of the same repeated syllable produced by different speakers with blocks of different syllables produced by the same voice. The right anterior STS was the only region that differed between these conditions. Here, in a similar passive task but with complex linguistic stimuli, we were not able to observe a similar adaptation effect. However, we might also have missed this region because of signal loss due to magnetic susceptibility artifacts (the peak maximum Talairach coordinates: 58, 2, −8, reported by Belin and Zatorre [2003] is outside the intersection of all the participants' masks). Von Kriegstein and Giraud [2004] also noticed that the right posterior STS was activated when participants had to recognize a precise voice, particularly if the voice was previously unknown. This finding is compatible with the exemplar-based representation of sentences that we observed in that region. In order to recognize a particular voice, the participants may emphasize the surface details encoded in this region.

CONCLUSION

Our results demonstrate the feasibility of using sentence-level adaptation to probe the functional organization of cortical language areas. Experiment 2 was a first step to segregate the perisylvian network by contrasting voice and linguistic content. Hierarchical comparisons between the different linguistic parameters should normally follow. A recent experiment [Kouider and Dupoux, 2005] demonstrating that it is possible to obtain subliminal priming with auditory stimuli opens the possibility of more refined tools to study speech comprehension, as was done for visual perception.

REFERENCES

Andoh J, Artiges E, Pallier C, Riviere D, Mangin JF, Cachia A, Plaze M, Paillere-Martinot ML, Martinot JL (2006): Modulation of language areas with functional MR image-guided magnetic stimulation. Neuroimage 29:619–627.

Belin P, Zatorre RJ (2003): Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14:2105–2109.

Belin P, Fecteau S, Bedard C (2004): Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8:129–135.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000): Human temporal lobe activation by speech and non speech sounds. Cereb Cortex 10:512–528.

Boemio A, Fromm S, Braun A, Poeppel D (2005): Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395.

Burton MW, Small SL, Blumstein SE (2000): The role of segmentation in phonological processing: an fMRI investigation. J Cogn Neurosci 12:679–690.

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

Caramazza A, Chialant D, Capasso R, Miceli G (2000): Separable processing of consonants and vowels. Nature 403:428–430.

Church BA, Schacter DL (1994): Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. J Exp Psychol Learn Mem Cogn 20:521–533.

Cohen L, Jobert A, Le Bihan D, Dehaene S (2004): Distinct unimodal and multimodal regions for word processing in the left temporal cortex. Neuroimage 23:1256–1270.

Davis MH, Johnsrude IS (2003): Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Dehaene S, Dupoux E, Mehler J, Cohen L, Paulesu E, Perani D, van de Moortele PF, Lehéricy S, Le Bihan D (1997): Anatomical variability in the cortical representation of first and second languages. Neuroreport 8:3809–3815.

Dehaene S, Jobert A, Naccache L, Ciuciu P, Poline JB, Le Bihan D, Cohen L (2004): Letter binding and invariant recognition of masked words: behavioral and neuroimaging evidence. Psychol Sci 15:307–313.

Dehaene-Lambertz G (2000): Cerebral specialization for speech and non-speech stimuli in infants. J Cogn Neurosci 12:449–460.

Dehaene-Lambertz G, Dehaene S. (1994): Speed and cerebral correlates of syllable discrimination in infants. Nature 370:292–295.

Dehaene-Lambertz G, Gliga T (2004): Common neural basis for phoneme processing in infants and adults. J Cogn Neurosci 16:1375–1387.

Dehaene-Lambertz G, Houston D (1998): Faster orientation latency toward native language in two-month-old infants. Lang Speech 41:21–43.

Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolle L, Dehaene S (2005): Neural correlates of switching from auditory to speech perception. Neuroimage 24:21–33.

Dronkers NF (1996): A new brain region for coordinating speech articulation. Nature 384:159–161.

Goldinger SD (1996): Words and voices: episodic traces in spoken word identification and recognition memory. J Exp Psychol Learn Mem Cogn 22:1166–1183.

Goldinger SD (1998): Echoes of echoes? An episodic theory of lexical access. Psychol Rev 105:251–279.

Grill-Spector K, Malach R (2001): fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst) 107:293–321.

Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R (1998): Cue-invariant activation in object-related areas of the human occipital lobe. Neuron 21:191–202.

Hahne A, Friederici AD (1999): Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. J Cogn Neurosci 11:194–205.

Hart J Jr, Berndt RS, Caramazza A (1985): Category-specific naming deficit following cerebral infarction. Nature 316:439–440.

Hashimoto R, Sakai KL (2002): Specialization in the left prefrontal cortex for sentence comprehension. Neuron 35:589–597.

Henson RN, Price CJ, Rugg MD, Turner R, Friston KJ (2002): Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15:83–97.

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

Kaan E, Swaab TY (2002): The brain circuitry of syntactic comprehension. Trends Cogn Sci 6:350–356.

Kisilevsky BS, Hains SM, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z (2003): Effects of experience on fetal voice recognition. Psychol Sci 14:220–224.

Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G (2002): Hearing sounds, understanding actions: action representation in mirror neurons. Science 297:846–848.

Kouider S, Dupoux E (2005): Subliminal speech priming. Psychol Sci 16:617–625.

Liberman AM, Mattingly IG (1985): The motor theory of speech perception revised. Cognition 21:1–36.

Luce PA, Lyons EA (1998): Specificity of memory representations for spoken words. Mem Cogn 26:708–715.

Mehler J, Bertoncini J, Barrière M, Jassik-Gerschenfeld D (1978): Infant recognition of mother's voice. Perception 7:491–497.

Mériaux S, Roche A, Dehaene-Lambertz G, Thirion B, Poline JB (2006): Combined permutation test and mixed-effect model for group average analysis in fMRI. Hum Brain Mapp 27:402–410.

Miller EK, Li L, Desimone R (1991): A neural mechanism for working and recognition memory in inferior temporal cortex. Science 254:1377–1379.

Naccache L, Dehaene S (2001): The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. Cereb Cortex 11:966–974.

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM (2003): Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362–1368.

Noppeney U, Price CJ (2004): An fMRI study of syntactic adaptation. J Cogn Neurosci 16:702–713.

Pallier C, Dehaene S, Poline JB, LeBihan D, Argenti AM, Dupoux E, Mehler J (2003): Brain imaging of language plasticity in adopted adults: can a second language replace the first? Cereb Cortex 13:155–161.

Petrides M, Pandya DN (2002): Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. Eur J Neurosci 16:291–310.

Piazza M, Izard V, Pinel P, Le Bihan D, Dehaene S (2004): Tuning curves for approximate numerosity in the human intraparietal sulcus. Neuron 44:547–555.

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

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999): Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2:1131–1136.

Sakai KL, Noguchi Y, Takeuchi T, Watanabe E (2002): Selective priming of syntactic processing by event-related transcranial magnetic stimulation of Broca's area. Neuron 35:1177–1182.

Schacter DL, Buckner RL, Koutstaal W, Dale AM, Rosen BR (1997): Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. Neuroimage 6:259–269.

Scott SK, Johnsrude IS (2003): The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100–107.

Thierry G, Boulanouar K, Kherif F, Ranjeva JP, Demonet JF (1999): Temporal sorting of neural components underlying phonological processing. Neuroreport 10:2599–2603.

Thierry G, Ibarrola D, Demonet JF, Cardebat D (2003): Demand on verbal working memory delays haemodynamic response in the inferior prefrontal cortex. Hum Brain Mapp 19:37–46.

Ulanovsky N, Las L, Nelken I (2003): Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391–398.

von Kriegstein K, Giraud AL (2004): Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948–955.

Vuilleumier P, Henson RN, Driver J, Dolan RJ (2002): Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat Neurosci 5:491–499.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992): Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.