

# Phonological Grammar Shapes the Auditory Cortex: A Functional Magnetic Resonance Imaging Study

Charlotte Jacquemot,<sup>1,2</sup> Christophe Pallier,<sup>3</sup> Denis LeBihan,<sup>3</sup> Stanislas Dehaene,<sup>3</sup> and Emmanuel Dupoux<sup>1</sup>

<sup>1</sup>Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole des Hautes Etudes en Sciences Sociales–Ecole Normale Supérieure, Centre National de la Recherche Scientifique, 75006 Paris, France, <sup>2</sup>Institut National de la Santé et de la Recherche Médicale, Université Paris XII, Hôpital Henri Mondor, 94010 Cedex Créteil, France, and <sup>3</sup>Institut National de la Santé et de la Recherche Médicale, Service Hospitalier Frédéric Joliot and Commissariat à l’Energie Atomique, 91401 Cedex Orsay, France

Languages differ depending on the set of basic sounds they use (the inventory of consonants and vowels) and on the way in which these sounds can be combined to make up words and phrases (phonological grammar). Previous research has shown that our inventory of consonants and vowels affects the way in which our brains decode foreign sounds (Goto, 1971; Näätänen et al., 1997; Kuhl, 2000). Here, we show that phonological grammar has an equally potent effect. We build on previous research, which shows that stimuli that are phonologically ungrammatical are assimilated to the closest grammatical form in the language (Dupoux et al., 1999). In a cross-linguistic design using French and Japanese participants and a fast event-related functional magnetic resonance imaging (fMRI) paradigm, we show that phonological grammar involves the left superior temporal and the left anterior supramarginal gyri, two regions previously associated with the processing of human vocal sounds.

**Key words:** speech perception; cross linguistic study; language specific illusion; planum temporale; fMRI; phonological grammar

## Introduction

Languages differ considerably depending on not only their inventory of consonants (Cs) and vowels (Vs) but also the phonological grammar that specifies how these sounds can be combined to form words and utterances (Kaye, 1989). Regarding the inventory of consonants and vowels, research has shown that infants become attuned to the particular sound categories used in their linguistic environment during the first year of life (Werker and Tees, 1984a; Kuhl et al., 1992). In adults, these categories strongly influence the way in which foreign sounds are perceived (Abramson and Lisker, 1970; Goto, 1971; Miyawaki et al., 1975; Trehub, 1976; Werker and Tees, 1984b; Kuhl, 1991), causing severe problems in the discrimination between certain non-native sounds. For instance, Japanese listeners have persistent trouble discriminating between English /r/ and /l/ (Goto, 1971; Lively et al., 1994). The current interpretation of these effects is that experience with native categories shapes the early acoustic–phonetic speech-decoding stage (Best and Strange, 1992; Best, 1995; Flege, 1995; Kuhl, 2000). Language experience has been found to modulate the mismatch negativity (MMN) response, which is supposed to originate in the auditory cortex (Kraus et al., 1995; Dehaene-Lambertz, 1997; Näätänen et al., 1997; Dehaene-Lambertz and Baillet, 1998; Sharma and Dorman, 2000).

Regarding phonological grammar, its role has been primarily studied by linguists, starting with early informal reports (Polivanov, 1931; Sapir, 1939) and more recently with the study of loanword adaptations (Silverman, 1992; Hyman, 1997). Although these studies do not include experimental tests, they suggest a strong effect of phonological grammar on perception. For instance, Japanese is primarily composed of simple syllables of the consonant–vowel type and does not allow complex strings of consonants, whereas English and French do allow these complex strings. Conversely, Japanese allows a distinction between short and long vowels, whereas English and French do not (e.g., “tokei” and “tookei” are two distinct words in Japanese). Accordingly, when Japanese speakers borrow foreign words, they insert so-called “epenthetic” vowels (usually /u/) into illegal consonant clusters so that the outcome fits the constraints of their grammar: the word “sphinx” becomes “sufinkusu” and the word “Christmas” becomes “Kurisumasu.” Conversely, when English or French import Japanese words, they neglect the vowel length distinction: “Tookyoo” becomes “Tokyo” and “Kyooto” becomes “Kyoto.” Recent investigations have claimed that such adaptations result from perceptual processes (Takagi and Mann, 1994; Dupoux et al., 1999, 2001; Dehaene-Lambertz et al., 2000). The current hypothesis is that the decoding of continuous speech into consonants and vowels is guided by the phonological grammar of the native language; illegal strings of consonants or vowels are corrected through insertion (Dupoux et al., 1999) or substitution of whole sounds (Massaro and Cohen, 1983; Halle et al., 1998). For instance, Dupoux et al. (1999) found that Japanese listeners have trouble distinguishing “ebza” from “ebuza,” and Dehaene-Lambertz et al. (2000) reported that this contrast does not generate a significant MMN, contrary to what is found with

Received June 10, 2003; revised Aug. 21, 2003; accepted Aug. 25, 2003.

This work was supported by a Cognitive PhD scholarship to C.J., an Action Incitative grant to C.P., a Cognitive grant, and a BioMed grant. We thank G. Dehaene-Lambertz, P. Ciuciu, E. Giacomini, N. Golestani, S. Franck, F. Hennen, J.-F. Mangin, S. Peperkamp, M. Peschanski, J.-B. Poline, and D. Rivière for help with this work.

Correspondence should be addressed to Charlotte Jacquemot, Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole des Hautes Etudes en Sciences Sociales, 54 bd Raspail, 75006 Paris, France. E-mail: jacquemot@lscp.ehess.fr.

Copyright © 2003 Society for Neuroscience 0270-6474/03/239541-06\$15.00/0

**Table 1.** Two examples of change trials and the condition to which they belong as a function of the native language of the participant

| Stimuli |        |       | French       | Japanese     |
|---------|--------|-------|--------------|--------------|
|         |        |       | Phonological | Acoustic     |
| ebza    | ebza   | ebuza |              |              |
|         |        |       | Acoustic     | Phonological |
| ebuuza  | ebuuza | ebuza |              |              |

The three auditory stimuli of each example are presented with their spectrogram.

French listeners. This suggests that the process that turns non-grammatical sequences of sounds into grammatical ones may take place at an early locus in acoustic–phonetic processing, probably within the auditory cortex.

In the present study, we aimed at identifying the brain regions involved in the application of phonological grammar during speech decoding. We built on previous studies to construct a fully crossed design with two populations (Japanese and French) and two contrasts (ebuza–ebuuza, and ebza–ebuza). One contrast, ebuza–ebuuza, is licensed by the phonological grammar of Japanese but not in French, in which differences in vowel length are not allowed within words. In French, both ebuza and ebuuza receive the same phonological representation (ebuza), and French participants can discriminate these stimuli only by relying on the acoustic differences between them. The other contrast, ebza–ebuza, has the same characteristics in reverse. It can be distinguished phonologically by the French participants but only acoustically by the Japanese participants. To make acoustic discrimination possible, we presented the contrasts without any phonetic variability, that is, the tokens were always spoken by the same speaker and, when identical, were physically identical. Indeed, previous research has found that phonetic variability increases the error rate for acoustic discriminations considerably (Werker and Tees, 1984b; Dupoux et al., 1997). Here, our aim was to obtain good performance on both acoustic and phonological discrimination but show that these two kinds of discrimination nonetheless involve different brain circuits.

French and Japanese volunteers were scanned while performing an AAX discrimination task. In each trial, three pseudowords were presented; the first two were always identical, and the third was either identical or different. When identical, all stimuli were acoustically the same. When different, the third item could differ from the other two in vowel duration (e.g., ebuza and ebuuza) or in the presence or absence of a vowel “u” (e.g., ebza and ebuza). As explained above, the change that was phonological for one population was only acoustic for the other (Table 1). Hence, by subtracting the activations involved in the phonological versus the acoustic discriminations, the brain areas that are involved in phonological processing alone can be pinpointed (Binder, 2000). Such a comparison is free of stimulus artifacts because across the two populations, the stimuli involved in the phonological and acoustic contrasts are exactly the same.

## Materials and Methods

**Participants.** Seven native speakers of Japanese 25–36 years of age (mean, 27) and seven native speakers of French 21–30 years of age (mean, 25) were recruited in Paris and participated in the study after giving written informed consent. All Japanese participants had started studying English after the age of 12 and French after the age of 18. None of the French participants studied Japanese. All participants were right-handed ac-

ording to the Edinburgh inventory. None had a history of neurological or psychiatric disease or hearing deficit.

**Stimuli and task.** The stimuli were the 20 triplets of pseudowords described by Dupoux et al. (1999). They followed the pattern VCCV/VCCV/VCCV (e.g., ebza/ebuza/ebuuza). For the present experiment, to present the three stimuli in the 2 sec silent window (see below), the stimuli were compressed to 60% of their original duration using the Psola algorithm in the Praat software (available at <http://www.praat.org>) so that their final duration was on average 312 msec ( $\pm 43$  msec). A fast event-related fMRI paradigm was used. Each trial lasted 3.3 sec and was composed of a silent window of 2 sec during which three stimuli were presented through headphones mounted with piezoelectric speakers (stimulus onset asynchrony; 600 msec), followed by 1.3 sec of fMRI acquisition. Thus, the noise of the gradients of the scanner did not interfere with the presentation of the stimuli. Trials were administered in sessions of 100, with each session lasting 6 min. Trials were of five types: acoustic change, acoustic no-change, phonological change, phonological no-change, and silence. The first four types corresponded to the crossing of two variables: acoustic versus phonological and change versus no-change. The acoustic versus phonological variable was defined as a function of the language of the subject (Table 1). The no-change trials contained the same items as the corresponding change trials, except that the three stimuli were physically identical. Within a session, 20 trials of each type were presented in random order. After performing a practice session, each participant performed between four and six experimental sessions during fMRI scanning.

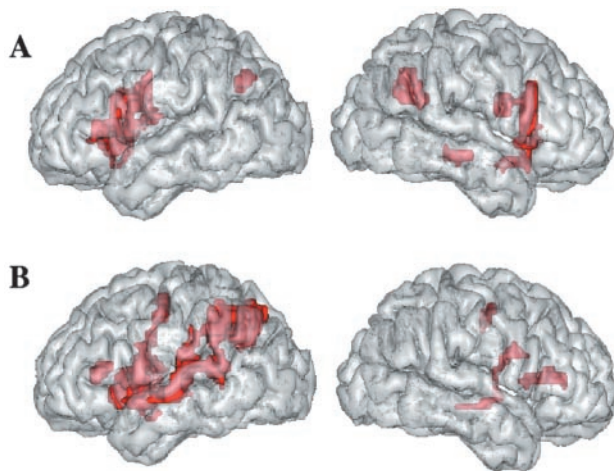
Participants were instructed that they would hear a series of three auditory stimuli, of which the first two would always be identical, and that they had to judge whether the last stimulus was strictly (physically) identical to the first two. They indicated their responses (same or different) by pressing a response button either with their left or right thumb. The response side was changed at midpoint during the experiment.

**Brain imaging.** The experiment was performed on a 3-T whole-body system (Bruker, Ettlingen, Germany) equipped with a quadrature bird-cage radio frequency coil and a head-gradient coil insert designed for echo planar imaging. Functional images were obtained with a T2-weighted gradient echo, echo planar-imaging sequence (repetition time, 3.3 sec; echo time, 40 msec; field of view,  $240 \times 240$  mm<sup>2</sup>; matrix,  $64 \times 64$ ). Each image, acquired in 1.3 sec, was made up of 22 4-mm-thick axial slices covering most of the brain. A high-resolution ( $1 \times 1 \times 1.2$  mm) anatomical image using a three-dimensional gradient-echo inversion-recovery sequence was also acquired for each participant.

fMRI data analysis was performed using statistical parametric mapping (SPM99; Welcome Department of Cognitive Neurology, London, UK). Preprocessing involved the following (in this order): slice timing, movement correction, spatial normalization, and smoothing (kernel, 5 mm). The resulting functional images had cubic voxels of  $4 \times 4 \times 4$  mm<sup>3</sup>. Temporal filters (high-pass cutoff at 80 sec; low-pass Gaussian width, 4 sec) were applied. For each participant, a linear model was generated by entering five distinct variables corresponding to the onsets of each of the five types of trials (acoustic change, acoustic no-change, phonological change, phonological no-change, and silence). Planned contrast images were obtained and then smoothed with a 6 mm Gaussian kernel and submitted to one-sample *t* tests (random effect analysis). Unless specified, the threshold for significance was set at  $p < 0.001$ , voxel-based uncorrected, and  $p < 0.05$ , corrected for spatial extent.

## Results

The analysis of the behavioral results revealed that the participants (combining Japanese and French groups) were globally able to detect the change in both conditions (90% correct). Reaction times and error rates were submitted to ANOVA with the factors language (Japanese vs French) and condition (phonological vs acoustic). The phonological condition was overall easier than the acoustic condition (error rates, 5.6 vs 13.6%;  $F_{(1,12)} = 25.1$ ;  $p < 0.001$ ; reaction times, 707 vs 732 msec;  $F_{(1,12)} = 5.9$ ;  $p < 0.05$  for the phonological vs acoustic condition, respectively). There were no main effects of language but in the analysis of



**Figure 1.** *A, B*, Activation rendered on the left hemisphere (left) and right hemisphere (right) of the brain. *A*, Areas activated by an acoustic change (reaching significance in the comparison of an acoustic change vs no-change conditions). *B*, Areas activated by a phonological change (reaching significance in the comparison of phonological change vs no-change conditions). Group analysis, voxel-based threshold at  $p < 0.001$ , uncorrected; spatial extent threshold,  $p < 0.05$ .

errors only, there was a significant interaction between language and condition ( $F_{(1,12)} = 14.4$ ;  $p < 0.01$ ). *Post hoc* comparisons of the errors showed that the effect of condition was significant in the Japanese (3.1 vs 17.2%;  $p < 0.01$ ; phonological vs acoustic condition, respectively) but not in the French group (8 vs 9.9%;  $p > 0.1$ ). *Post hoc* comparisons of the reaction times showed that the effect of condition was significant in the French (690 vs 725 msec;  $p < 0.05$ ; phonological vs acoustic condition, respectively) but not in the Japanese group (724 vs 739 msec;  $p > 0.1$ ). Such an asymmetry between speed and accuracy across languages was already observed by Dupoux et al. (1999) but in both languages, the conclusion is the same: there is an advantage for the phonological condition relative to the acoustic condition. The overall size of the phonological effect is smaller than in the Dupoux et al. study, because we purposefully used a situation with only one speaker voice to facilitate the discrimination on the basis of acoustic differences.

In analyzing the fMRI data, we computed three contrasts, one to identify the circuits involved in the detection of an acoustic change, one for the circuits involved in the detection of a phonological change, and one for the difference between the two circuits.

First, we calculated contrast images between the acoustic change and acoustic no-change conditions. Acoustic change activated a large network, comprising the right superior and middle temporal gyri and, bilaterally, the intraparietal sulci, inferior frontal gyri, insula, cingulate cortex, and thalamus (Fig. 1*A*; Table 2), which is congruent with previous studies (Zatorre et al., 1994; Belin et al., 1998). Second, we calculated the contrast between the phonological change and phonological no-change conditions. Phonological change caused activation in the perisylvian areas in the left hemisphere, including the inferior frontal gyrus, superior temporal gyrus (STG), supramarginal and angular gyri, and left intraparietal sulcus (Fig. 1*B*; Table 3), typically associated with discrimination tasks involving speech sound analysis (Démonet et al., 1992; Zatorre et al., 1992; Burton et al., 2000). Significant activation was also observed bilaterally in the cingulate cortex, insula, and precentral gyrus. To a lesser extent, the right inferior frontal and the right superior and middle temporal gyri

**Table 2. Brain areas activated by the detection of an acoustic change**

| Brain area  | z-score | Peak location in Talairach coordinates (x, y, z; mm) |
|---|---------|--|
| L inferior frontal gyrus, pars triangularis BA 45 | 3.85    | −44, 28, 4   |
| L inferior frontal gyrus, pars opercularis BA 44  | 3.54    | −44, 12, 8   |
| L inferior frontal gyrus BA 44/45                 | 3.60    | −44, 12, 20  |
| L intraparietal sulcus BA 39/40                   | 4.01    | −32, −64, 40   |
| L thalamus  | 3.99    | −16, −4, 16  |
| L central sulcus BA 3/4                           | 3.93    | −36, 0, 28   |
| L insula  | 3.80    | −28, 20, 4   |
| R inferior frontal gyrus, pars opercularis BA 44  | 4.02    | 48, 16, 8  |
| R inferior frontal gyrus BA 44/45                 | 3.73    | 48, 16, 24   |
| R superior temporal sulcus (anterior) BA 21/22    | 3.92    | 48, 4, −16   |
| R middle temporal gyrus BA 21                     | 3.87    | 56, −28, −8  |
| R intraparietal sulcus BA 39/40                   | 4.60    | 28, −48, 24  |
| R central sulcus BA 3/4                           | 3.52    | 40, 0, 20  |
| R insula  | 4.33    | 24, 24, 4  |
| R thalamus  | 4.14    | 12, 0, 12  |
| Cingulate sulcus BA 32/8/9                        | 3.97    | −4, 24, 36   |
| Cingulate sulcus BA 32/9                          | 3.95    | 4, 24, 32  |

Coordinates, in standard stereotactic space of Talairach and Tournoux (Talairach and Tournoux, 1988), refer to maxima of the z value within each focus of activation. L, Left; R, right. Approximate Brodmann numbers (BA) associated with anatomical regions are given.

Group analysis: threshold set at  $p < 0.001$ , uncorrected, and  $p < 0.05$ , corrected for spatial extent.

**Table 3. Brain areas activated by the detection of a phonological change**

| Brain area  | z score | Peak location in Talairach coordinates (x, y, z; mm) |
|---|---------|--|
| L intraparietal sulcus BA 31/7                    | 5.23    | −32, −60, 36   |
| L supramarginal gyrus (anterior) BA 40            | 3.91    | −56, −28, 28   |
| L supramarginal gyrus (posterior) BA 40           | 3.81    | −44, −40, 40   |
| L angular gyrus BA 39                             | 3.81    | −52, −44, 48   |
| L superior temporal sulcus BA 21/22               | 3.37    | −48, −8, −16   |
| L superior temporal sulcus (posterior) BA 21/22   | 4.22    | −48, −48, 8  |
| L superior temporal gyrus (posterior) BA 22/40    | 4.45    | −56, −44, 12   |
| L superior temporal gyrus BA 22/42                | 4.35    | −64, −16, 0  |
| L inferior frontal gyrus, pars triangularis BA 45 | 3.84    | −48, 32, 8   |
| L inferior frontal gyrus, pars opercularis BA 44  | 4.23    | −52, 8, 12   |
| L insula  | 4.15    | −40, −8, 12  |
| L precentral gyrus BA 4/6                         | 4.04    | −36, −8, 56  |
| R inferior frontal gyrus, pars triangularis BA 45 | 3.86    | 40, 28, 12   |
| R middle temporal gyrus BA 21/22                  | 3.34    | 60, −24, −8  |
| R superior temporal gyrus BA 22                   | 3.36    | 60, −12, −8  |
| R central sulcus BA 3/4                           | 3.53    | 24, −12, 44  |
| R precentral gyrus BA 6/9                         | 3.85    | 40, 8, 28  |
| R insula  | 3.62    | 40, 16, 8  |
| R lingual gyrus BA 17/18                          | 3.41    | 8, −76, 4  |
| Cingulate sulcus BA 32/8                          | 4.32    | −4, 8, 52  |
| Cingulate sulcus BA 32/8                          | 3.63    | 8, 12, 40  |

Coordinates, in standard stereotactic space of Talairach and Tournoux (Talairach and Tournoux, 1988), refer to maxima of the z value within each focus of activation. L, Left; R, right. Approximate Brodmann numbers (BA) associated with anatomical regions are given.

Group analysis: threshold set at  $p < 0.001$ , uncorrected, and  $p < 0.05$ , corrected for spatial extent.

were also activated. Regions activated in both conditions were the insula, cingulate cortex, and central sulcus. These regions have been shown to be involved in the motor and cognitive components of an auditory task requiring attention and motor response (Zatorre et al., 1994). Finally, we calculated the difference between the phonological and acoustic change circuits. We found two regions that were significantly more activated by the phonological than the acoustic changes: the left STG and anterior part of the left supramarginal gyrus (SMG) (Fig. 2). When the threshold was lowered to  $p < 0.01$ , a region in the right STG also appeared

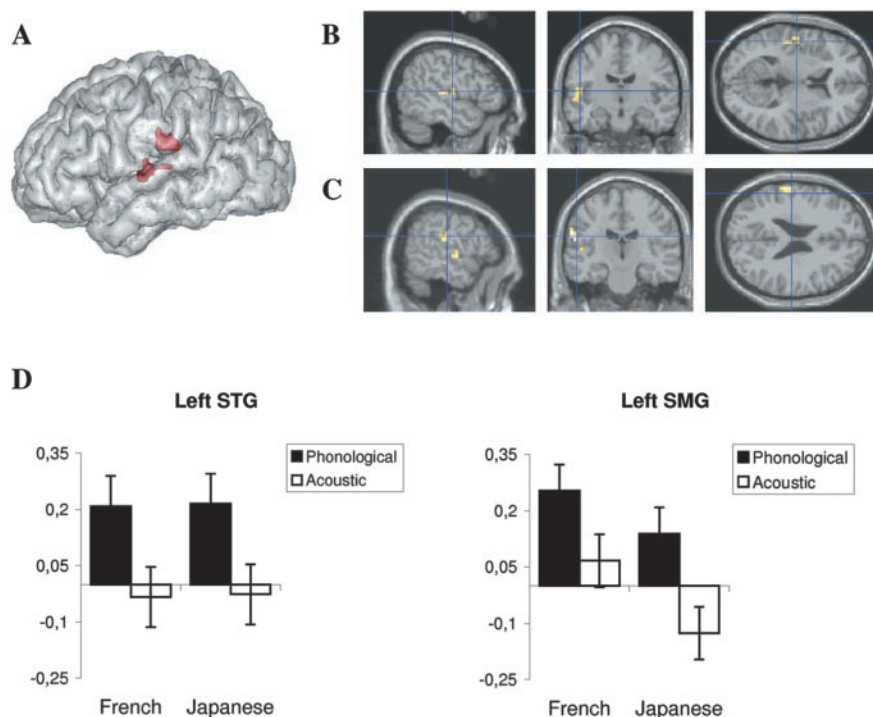


( $x = 52$ ;  $y = -8$ ;  $z = 4$ ;  $z$ -score, 3.4; cluster size, 71;  $p = 0.036$ , corrected). No region was significantly more activated by the acoustic changes than by the phonological changes.

## Discussion

We found a phonological grammar effect in two regions in the left hemisphere: one in the STG and one located in the anterior SMG (Fig. 2). There was more activation in these regions when the stimuli changed phonologically than when they changed acoustically. These activations were found by comparing the same two sets of stimuli across French and Japanese speakers. In principle, participants could discriminate against all stimuli solely on the basis of acoustic features. However, our results suggest that a phonological representation of the stimuli was activated and informed the discrimination decision. This is confirmed by behavioral data that show performance was slightly but significantly better in the phonological condition than in the acoustic condition.

The peak activation in the left STG lies on the boundary between the Heschl gyrus (HG) and the planum temporale (PT). Atlases (Westbury et al., 1999; Rademacher et al., 2001) indicate ~40–60% of probability of localization in either structure (note that the activation observed in the right STG when lowering the statistical threshold is probably located in the Heschl gyrus). Because it is generally believed that the PT handles more complex computations than the primary auditory cortex (Griffiths and Warren, 2002), it is reasonable to think that the complex process of phonological decoding takes place in the PT. Yet, the current state of knowledge does not allow to categorically claim that HG cannot support this process. Jäncke et al. (2002) observed activations that also straddled the PT and HG when comparing unvoiced versus voiced consonants. Numerous studies have revealed increases of PT activations with the spectrotemporal complexity of sounds (for review, see Griffiths and Warren, 2002; Scott and Johnsrude, 2003). The present data indicate that PT activations do not simply depend on the acoustic complexity of speech sounds but also reflect processes tuned to the phonology of the native language. This result adds to the converging evidence in favor of the involvement of the PT in phonological processing. First, lesions in this region can provoke word deafness, the inability to process speech sounds with hearing acuity within normal limits (Metz-Lutz and Dahl, 1984; Otsuki et al., 1998), and syllable discrimination can be disrupted by electrical interference in the left STG (Boatman et al., 1995). Second, activity in the PT has been observed in lip-reading versus watching meaningless facial movements (Calvert et al., 1997) when profoundly deaf signers process meaningless parts of signs corresponding to syllabic units (Petitto et al., 2000) and when reading (Nakada et al., 2001). Finally, PT activations have also been reported in speech production (Paus et al., 1996). These data are consistent with the notion that the PT subserves the computation of an amodal, abstract, phonological representation.



**Figure 2.** Areas significantly more activated by a phonological change than by an acoustic change. *A*, Rendering on a three-dimensional left hemisphere template. *B*, *C*, Sections centered on the two local maxima in the left STG (*B*) (coordinates in standard stereotaxic space of Talairach and Tournoux:  $x = -48$  mm;  $y = -24$  mm;  $z = 8$  mm;  $z$ -score, 3.65; cluster size, 14 voxels) and in the left SMG (*C*) (coordinates:  $x = -60$  mm;  $y = -20$  mm;  $z = 28$  mm;  $z$ -score, 3.92; cluster size, 17 voxels). *D*, Plots of the size of the effect at the two local maxima, as a function of condition and language (Japanese and French). Scale bars show the mean percentage signal change ( $\pm$  SE) for each of the following conditions: phonological (phonological change vs phonological no-change) and acoustic (acoustic change vs acoustic no-change).

The second region activated by phonological change was located in the left SMG. Focal lesions in this region are not typically associated with auditory comprehension deficits (Hickok and Poeppel, 2000) and are not reported when people listen to speech (Crinion et al., 2003). Yet activations in the SMG have been observed when subjects had to perform experimental tasks involving phonological short-term memory (Paulesu et al., 1993; Celsus et al., 1999). A correlation and regression analysis has also revealed that patients impaired in syllable discrimination tend to have lesions involving the left SMG (Caplan et al., 1995). Thus, the left SMG activation found in the present study may be linked to working memory processes and processes translating from auditory to articulatory representations that can be involved in speech discrimination tasks (Hickok and Poeppel, 2000).

Remarkably, we did not find that frontal areas were more involved in the phonological condition than in the acoustic condition, even when the threshold was lowered. This differs from neuroimaging studies that have claimed that phonological processing relies on left inferior frontal regions (Demonet et al., 1992; Zatorre et al., 1992; Hsieh et al., 2001; Gandour et al., 2002). These studies have used tasks that require the explicit extraction of an abstract linguistic feature, such as phoneme, tone, or vowel duration. Such explicit tasks are known to depend on literacy and engage orthographic representations (Morais et al., 1986; Poeppel, 1996). Burton et al. (2000) claimed that frontal activation is found only in tasks that require explicit segmentation into consonants and vowels and those that place high demands on working memory. In the present study, the task does not require segmentation of the auditory stimuli.

Previous research on speech processing has focused on the

effects of consonant and vowel categories. These categories are acquired early by preverbal infants (Werker and Tees, 1984a; Kuhl et al., 1992; Maye et al., 2002), affect the decoding of speech sounds (Goto, 1971; Werker and Tees, 1984b), and involve areas of the auditory cortex (Näätänen et al., 1997; Dehaene-Lambertz and Baillet, 1998). In contrast, the effect of phonological grammar has been less studied but also seems to be acquired early (Jusczyk et al., 1993, 1994) and shapes the decoding of speech sounds (Massaro and Cohen, 1983; Dupoux et al., 1999; Dehaene-Lambertz et al., 2000). At first sight, the regions we found (left STG and SMG) might be the same as those involved in consonant and vowel processing. Additional research is needed to establish whether these regions uniformly represent the different aspects of the sound system, or whether separate subparts of the STG sustain the processing of consonant and vowel categories on the one hand and phonological grammar on the other. This, in turn, could help us tease apart theories of perception that posit two distinct processing stages involving either phoneme identification or grammatical parsing (Church, 1987) from theories in which these two processes are merged into a single step of syllabic template matching (Mehler et al., 1990).

## References

- Abramson AS, Lisker L (1970) Discriminability along the voicing continuum: cross-language tests. In: Proceedings of the sixth international congress of phonetic sciences, pp 569–573. Prague: Academia.
- Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y (1998) Lateralization of speech and auditory temporal processing. *J Cogn Neurosci* 10:536–540.
- Best C, Strange W (1992) Effects of phonological and phonetic factors on cross-language perception of approximants. *J Phonetics* 20:305–331.
- Best CT (1995) Second-language speech learning: theory, findings, and problems. In: *Speech perception and linguistic experience: theoretical and methodological issues* (Strange W, Jenkins JJ, eds), pp 171–206. Timonium, MD: York.
- Binder J (2000) The new neuroanatomy of speech perception. *Brain* 123:2400–2406.
- Boatman D, Lesser RP, Gordon B (1995) Auditory speech processing in the left temporal lobe: an electrical interference study. *Brain Lang* 51:269–290.
- Burton MW, Small S, Blumstein SE (2000) The role of segmentation in phonological processing: an fMRI investigation. *J Cogn Neurosci* 12:679–690.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593–596.
- Caplan D, Gow D, Makris N (1995) Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology* 45:293–298.
- Celsis P, Boulouaou K, Doyon B, Ranjeva JP, 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.
- Church KW (1987) Phonological parsing and lexical retrieval. *Cognition* 25:53–69.
- Crinin JT, Lambon-Ralph MA, Warburton EA, Howard D, Wise RJ (2003) Temporal lobe regions engaged during normal speech comprehension. *Brain* 126:1193–1201.
- Dehaene-Lambertz G (1997) Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8:919–924.
- 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* 12:635–647.
- Démonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulous JL, Wise R, Rascol A, Frackowiak R (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain* 115:1753–1768.
- Dupoux E, Pallier C, Sebastian-Gallés N, Mehler J (1997) A “deafness” in French? *J Mem Lang* 36:406–421.
- Dupoux E, Kakehi K, Hirose Y, Pallier C, Fitneva S, Mehler J (1999) Epenthetic vowels in Japanese: a perceptual illusion. *J Exp Psychol Hum Percept Perform* 25:1568–1578.
- Dupoux E, Pallier C, Kakehi K, Mehler J (2001) New evidence for prelexical phonological processing in word recognition. *Lang Cogn Proc* 16:491–505.
- Flege J (1995) Second-language speech learning: theory, findings, and problems. In: *Speech perception and linguistic experience: theoretical and methodological issues* (Strange W, Jenkins JJ, eds), pp 233–273. Timonium, MD: York.
- Gandour J, Wong D, Lowe M, Dziedzic M, Saththamnuwong N, Tong Y, Li X (2002) A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *J Cogn Neurosci* 14:1076–1087.
- Goto H (1971) Auditory perception by normal Japanese adults of the sounds “l” and “r”. *Neuropsychologia* 9:317–323.
- Griffiths TD, Warren JD (2002) The planum temporale as a computational hub. *Trends Neurosci* 25:348–353.
- Halle PA, Segui J, Frauenfelder U, Meunier C (1998) Processing of illegal consonant clusters: a case of perceptual assimilation? *J Exp Psychol* 4:592–608.
- Hickok G, Poeppel D (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4:131–138.
- Hsieh L, Gandour J, Wong D, Hutchins GD (2001) Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang* 3:227–252.
- Hyman L (1997) The role of borrowings in the justification of phonological grammars. *Stud Afr Ling* 1:1–48.
- Jäncke L, Wustenberg T, Scheich H, Heinze HJ (2002) Phonetic perception and the temporal cortex. *NeuroImage* 4:733–746.
- Jusczyk PW, Friederici AD, Wessels JMI, Svenkerud VY, Jusczyk AM (1993) Infants’ sensitivity to the sound pattern of native language words. *J Mem Lang* 32:402–420.
- Jusczyk PW, Luce PA, Charles-Luce J (1994) Infants’ sensitivity to phonotactic patterns in the native language. *J Mem Lang* 33:630–645.
- Kaye JD (1989) *Phonology: a cognitive view*. Hillsdale, NJ: LEA.
- Kraus N, McGee T, Carrell T, King C, Tremblay K (1995) Central auditory system plasticity associated with speech discrimination training. *J Cogn Neurosci* 7:27–34.
- Kuhl PK (1991) Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories monkeys do not. *Percept Psychophys* 50:93–107.
- Kuhl PK (2000) A new view of language acquisition. *Proc Natl Acad Sci USA* 97:11850–11857.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B (1992) Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science* 255:606–608.
- Lively SE, Pisoni DB, Yamada RA, Tohkura Yi, Yamada T (1994) Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *J Acoust Soc Am* 96:2076–2087.
- Massaro DW, Cohen MM (1983) Phonological constraints in speech perception. *Percept Psychophys* 34:338–348.
- Maye J, Werker JF, Gerken L (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 3:101–111.
- Mehler J, Dupoux E, Segui J (1990) Constraining models of lexical access: the onset of word recognition. In: *Cognitive models of speech processing: psycholinguistic and computational perspectives* (Altmann G, ed), pp 236–262. Cambridge, MA: MIT.
- Metz-Lutz MN, Dahl E (1984) Analysis of word comprehension in a case of pure word deafness. *Brain Lang* 1:13–25.
- Miyawaki K, Strange W, Verbrugge R, Liberman AM, Jenkins JJ, Fujimura O (1975) An effect of linguistic experience: the discrimination of /r/ and /l/ by native speakers of Japanese and English. *Percept Psychophys* 18:331–340.
- Morais J, Bertelson P, Cary L, Alegria J (1986) Literacy training and speech segmentation. *Cognition* 24:45–64.
- Näätänen R, Lehtokovski A, Lenne M, Cheour M, Huottilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.
- Nakada T, Fujii Y, Yoneoka Y, Kwee IL (2001) Planum temporale: where spoken and written language meet. *Eur Neurol* 46:121–125.

- Otsuki M, Soma Y, Sato M, Homma A, Tsuji S (1998) Slowly progressive pure word deafness. *Eur Neurol* 39:135–140.
- Paulesu E, Frith CD, Frackowiak RSJ (1993) The neural correlates of the verbal component of working memory. *Nature* 362:342–345.
- Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236–2246.
- Petitto LA, Zatorre RJ, Gauna K, Nikelski EJ, Dostie D, Evans AC (2000) Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proc Natl Acad Sci USA* 97:13476–13477.
- Poeppel D (1996) A critical review of PET studies of phonological processing. *Brain Lang* 55:317–351.
- Polivanov E (1931) La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4:79–96.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HF, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* 13:669–683.
- Sapir E (1939) *Language: an introduction of study of speech*. New York: Harcourt Brace Jovanovich.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Sharma A, Dorman M (2000) Neurophysiologic correlates of cross-languages phonetic perception. *J Acoust Soc Am* 105:2697–2703.
- Silverman D (1992) Multiple scansions in loanword phonology: evidence from Kantonese. *Phonology* 9:289–328.
- Takagi N, Mann V (1994) A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *J Phonetics* 22:343–356.
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Trehub SE (1976) The discrimination of foreign speech contrasts by infants and adults. *Child Dev* 47:466–472.
- Werker JF, Tees RC (1984a) Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7:49–63.
- Werker JF, Tees RC (1984b) Phonemic and phonetic factors in adult cross-language speech perception. *J Acoust Soc Am* 75:1866–1878.
- Westbury CF, Zatorre RJ, Evans AC (1999) Quantifying variability in the planum temporale: a probability map. *Cereb Cortex* 9:392–405.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849.
- Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.