

Learning new sounds of speech: reallocation of neural substrates

Narly Golestani* and Robert J. Zatorre

Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University, Montreal, Canada

Received 18 February 2003; revised 5 August 2003; accepted 12 September 2003

Functional magnetic resonance imaging (fMRI) was used to investigate changes in brain activity related to phonetic learning. Ten monolingual English-speaking subjects were scanned while performing an identification task both before and after five sessions of training with a Hindi dental–retroflex nonnative contrast. Behaviorally, training resulted in an improvement in the ability to identify the nonnative contrast. Imaging results suggest that the successful learning of a nonnative phonetic contrast results in the recruitment of the same areas that are involved during the processing of native contrasts, including the left superior temporal gyrus, insula–frontal operculum, and inferior frontal gyrus. Additionally, results of correlational analyses between behavioral improvement and the blood-oxygenation-level-dependent (BOLD) signal obtained during the posttraining Hindi task suggest that the degree of success in learning is accompanied by more efficient neural processing in classical frontal speech regions, and by a reduction of deactivation relative to a noise baseline condition in left parieto-temporal speech regions.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Sounds of speech; Reallocation of neural substrates; Functional magnetic resonance imaging

Introduction

Infants aged 6 months or younger are able to discriminate speech sounds, including many that are not used to distinguish words in their native language. However, during development and starting as early as at 6 months of age, lack of experience with certain nonnative speech sounds results in a developmental shift from a language-general to a language-specific pattern of phonetic perception (Best et al., 1988; Jusczyk, 1995; Kuhl, 2000; Kuhl et al., 1992; Polka and Werker, 1994; Werker and Lalonde, 1988; Werker and Tees, 1984a). Most adults can better distinguish two speech sounds belonging to different phonetic categories than ones belonging to the same category, even when the physical differences separating the stimuli have been equated (Flege, 1984; Liberman, 1957; Liberman et al., 1957, 1967; Pisoni et al., 1982; Werker and Tees, 1984b). Despite native-language phonetic perception, adults are capable of learning new languages, and thereby of learning to

distinguish nonnative phonetic contrasts. Interestingly, even amongst adults with very similar language backgrounds, considerable individual differences exist in their ability to improve following phonetic training (Polka, 1991; Priutt et al., 1990; Strange and Dittman, 1984; Strange et al., 1989; Werker et al., 1981). This finding leads to important questions regarding the functional neural substrates underlying the perception of native versus newly learned, nonnative speech sounds, and more specifically, regarding possible differences in functional anatomy between individuals who successfully learn new speech sounds and those who do not benefit from training.

The neural correlates of phonetic perception have been studied using functional brain imaging techniques such as PET and fMRI. These experiments have involved auditory presentation of stimuli including words, speech syllables, and meaningless speech sounds, and tasks used have included passive listening, phoneme monitoring, discrimination, or identification, and rhyming judgments. Generally, the results have shown the involvement of regions in and around what is classically known as “Wernicke’s area”, including left-sided activations in perisylvian temporoparietal areas including the supramarginal and angular gyri (Binder et al., 1996, 1997; Démonet et al., 1994a; Paulesu et al., 1993; Petersen et al., 1988; Zatorre et al., 1992, 1996). Consistent with functional imaging work, there is also evidence from lesion studies that deficits in phonological processing may arise from damage to perisylvian regions in and around Wernicke’s area, including the left superior temporal gyrus and the supramarginal gyrus (Benson, 1967; Benson et al., 1973; Geschwind, 1970, 1971). Results of functional imaging work specifically examining phonetic perception have also typically shown activity in the superior temporal gyrus (STG) bilaterally (Binder et al., 1994; Hickok and Poeppel, 2000; Jäncke et al., 1998; Mazoyer et al., 1993; Mummery et al., 1999).

The involvement of regions in and around the frontal speech area classically known as Broca’s area in phonological processing has been the subject of controversy. Results of some studies involving receptive speech-related tasks have not shown activation in this region (Petersen et al., 1989; Rumsey et al., 1992). In contrast, a larger number of studies have shown its involvement in purely receptive language tasks that make certain specific demands (Burton et al., 2000; Démonet et al., 1992, 1994b; Fiez et al., 1995; Zatorre et al., 1992, 1996). The frontal regions identified differ across studies, making the interpretation of the roles of such regions more difficult. Although speech perception has not been investigated extensively in aphasic patients with lesions in and around Broca’s area, existing studies have shown deficits in

* Corresponding author. Service Hospitalier Frédéric Joliot, 4 Place du Général Leclerc, Orsay 91401, France. Fax: +33-1-69-86-78-16.

E-mail address: narlyg@bic.mni.mcgill.ca (N. Golestani).

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

phonetic discrimination (Blumstein et al., 1977; Tallal and Newcombe, 1978), and in temporal perception (Tallal and Newcombe, 1978) in such patients.

Plasticity of auditory function resulting from training and experience has been shown using techniques such as single cell recordings in animals (Kraus and Disterhoft, 1982; Recanzone et al., 1993), magnetoencephalography (Pantev et al., 1999), and event-related potentials (ERP) (Kraus et al., 1995; Tremblay et al., 1998) in humans. For example, behavioral training of two slightly different native speech stimuli in adults results in a significant change in the duration and magnitude of the mismatch negativity (MMN) (Kraus et al., 1995), an auditory cortical response to acoustic change that is introduced in a repetitive stimulus sequence (Näätänen et al., 1978, 1993). This physiological change precedes behavioral discrimination improvements (Tremblay et al., 1998), suggesting that the MMN is a measure of preattentive learning (see Kraus and Cheour, 2000). A number of studies show hemispheric asymmetries in the MMN (Alho et al., 1998; Csépe, 1995; Tervaniemi et al., 2000). Tremblay et al. (1997) showed that MMNs elicited by nonnative speech syllables were initially symmetrical, but that they became enhanced over the left hemisphere following training.

The aim of the present study was to determine how the pattern of brain activity may change as a result of training with speech sounds from a nonnative language. Subjects were scanned using fMRI before and after a 2-week period of phonetic training with a Hindi dental–retroflex contrast. During scanning, a native phonetic contrast was used as a control. A noise control condition was also used to subtract out lower level acoustic processing, and to make the results more comparable to those of previous studies on phonetic processing (Binder et al., 2000; Zatorre et al., 1992). We wanted to address the following questions. First, does the identification of newly learned speech sounds recruit the same neural substrates as does the identification of a known, native phonetic contrast, or are new areas recruited? The second question relates to whether we can differentiate “learners” from “non-learners” on the basis of their pattern of activation while they classify the new speech sounds. We predicted firstly that the native identification task would reveal the bilateral involvement of superior temporal regions, stronger in the left than in the right hemisphere, of the left temporoparietal region, and of the left inferior frontal gyrus (IFG) in and adjacent to Broca’s area. Second, based on the above reported lateralization of the MMN response to nonnative speech sounds following training, we predicted that before training, the neural response to nonnative speech sounds would be bilateral, but that it would be more left lateralized after training. We also predicted that after training, the pattern of activation outside of the auditory regions (i.e., in the left temporoparietal and inferior frontal regions) would be similar to that found in the native condition. This prediction is also based on results of neuroimaging studies of language function in healthy bilinguals, some of which show that at the single word level, brain regions subserving the native language (L1) and the second language (L2) in fluent bilinguals appear to overlap (Chee et al., 1999; Illes et al., 1999; Klein et al., 1994, 1995). Last, based on the assumption that more successful task performance recruits underlying neural substrates more actively, we predict that correlations between a behavioral learning measure and the blood-oxygenation-level-dependent (BOLD) signal during the posttraining nonnative task would reveal a positive relationship between learning and signal in left prefrontal and left temporoparietal speech areas.

Materials and methods

Subjects

Ten right-handed monolingual English-speaking participants (4 men), ranging in age from 20 to 29 participated in the study. None had been exposed to or had experience with languages in which the retroflex speech sound is phonologically represented.

Stimuli

Stimulus selection

We selected the dental–retroflex place-of-articulation contrast, which is used in languages of India such as Hindi or Urdu. Retroflex consonants require a relatively complex articulation, they are rare across languages (only 11% of the world’s languages include a retroflex consonant), and are not used phonemically in English. The nonnative perception of the retroflex sound is diminished due to perceptual reorganization early in development, below the age of one (Tees and Werker, 1984; Werker and Tees, 1983, 1984a; see Burnham, 1986), and perceptually, English listeners assimilate the dental–retroflex sounds such that they perceive both sounds as instances of the dental consonant (Polka, 1991; Werker and Lalonde, 1988).

Research aimed at training the nonnative perception of the dental and retroflex sounds in adults has shown that standard training approaches fail to improve performance (Tees and Werker, 1984; Werker and Tees, 1983, 1984a; Werker et al., 1981). Improvement has been reported under certain circumstances, including the use of shorter interstimulus intervals (ISI) (Werker and Logan, 1985; Werker and Tees, 1984b), truncated stimuli (Priutt et al., 1990), and extensive experience with the contrast (Tees and Werker, 1984).

We selected this nonnative retroflex consonant because we wanted to ensure that none of our participants had had phonemic exposure to this sound. Additionally, we wanted to ensure that listeners would have trouble distinguishing this sound from the dental consonant before training (to avoid pretraining ceiling effects), and that we would find a considerable range in improvement across subjects so that we could then examine brain activity as a function of success in learning. In a previous study (Golestani et al., submitted for publication), we synthesized the retroflex consonant, and showed firstly that phonetic training using a perceptual fading technique resulted in overall improvement in the ability to identify this sound versus the native dental one in a large group of subjects, and secondly, that there were considerable individual differences in the amount of learning across subjects.

Stimulus synthesis

Synthesis of the nonnative phonetic contrast was based on parameters reported by Stevens and Blumstein (1975) for the synthetic retroflex–dental place-of-articulation continuum. Refer to Fig. 1 for schematized frequency–time representations of the dental and retroflex sounds. There were seven stimuli varying in equal steps in terms of acoustic difference between adjacent items. The set of four-formant stimuli was constructed with the use of the MITSYN (Henke, 1990) signal processing language software and the Klatt model synthesizer. Stimulus 1 corresponds to the dental and stimulus 7 to the retroflex voiced, unaspirated stop consonant prototype, and all of the consonants were followed by the vowel /a/.

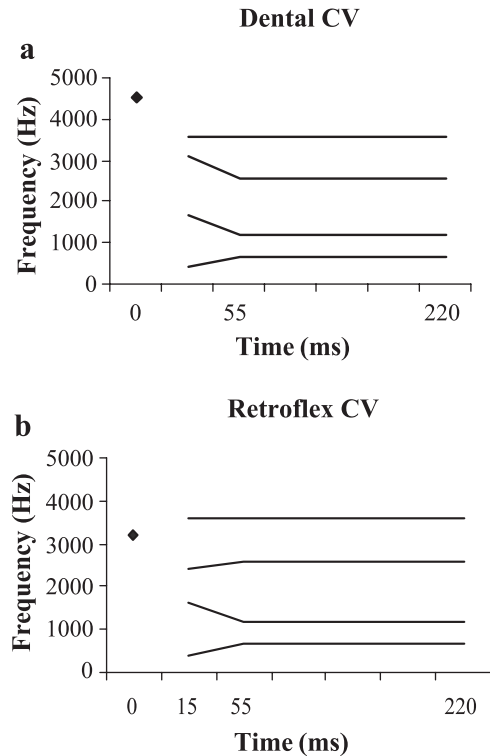


Fig. 1. Schematized frequency versus time diagrams of the (a) dental and (b) retroflex consonant-vowel (CV) syllables. Each line represents the center value of a formant trajectory; diamond symbol represents an onset noise burst.

Stimuli began with an initial noise burst. The parameters that were manipulated to create the continuum are the frequency glides of the third formant (F3), as well as the center frequency of the burst. The initial noise burst lasted 5 ms, and its center frequency decreased in equal step sizes of 217 Hz from 4500 (stimulus 1) to 3198 Hz (stimulus 7). The voicing began 15 ms after the onset of the burst. Each stimulus lasted 220 ms in total, and formant transitions (FTs) lasted 40 ms. The fundamental frequency (F0) started at 130 Hz and decreased linearly over the first 40 ms to 125 Hz, then remained steady-state at 125 Hz for 75 ms, and then gradually decreased to 90 Hz over the last 90 ms of the sound. The first formant (F1) rose from 400 to 655 Hz during the transition, and then remained steady-state during the remainder of the sound. The second formant (F2) decreased from 1650 to 1185 Hz over the first 40 ms, and then remained steady-state. The fourth formant (F4) remained steady-state at 3600 Hz throughout the entire sound. The starting frequency of the third formant (F3) varied in equal 111 Hz steps from 3080 (stimulus 1) to 2414 Hz (stimulus 7), and the steady-state portion of this formant was always at 2585 Hz.

The native control stimulus contrast was constituted of the dental /da/ sound described above (stimulus 1), and a synthetic voiceless stop consonant /t/, followed by the vowel /a/. (/ta/). Stimulus characteristics for the latter are the following. The stimulus lasted 220 ms, and the FTs lasted 30 ms. The initial noise burst lasted 10 ms, and its center frequency was 4000 Hz. Voicing began 50 ms after the onset of the burst. The F0 started at 130 Hz, decreased linearly for 40 ms to 125 Hz, remained steady-state for 40 ms, and then decreased linearly to 90 Hz over the last

90 ms of the sound. Formants 1, 2, and 3 remained steady-state after the FT. F1 rose from 530 to 655 Hz during the FT, F2 decreased from 1480 to 1185 Hz, and F3 decreased from 2857 to 2585 Hz. F4 remained steady-state at 3600 Hz during the entire stimulus.

The noise stimulus consisted of amplitude modulated white noise matched in sound pressure level (SPL) and duration with the CV stimuli. It was generated by using the same wave envelope as the speech stimuli. The interstimulus interval for the noise condition was also matched with that of the phonetic condition. Stimuli were presented binaurally with electrostatic, MR-compatible headphones (Koss) at an 87 dB SPL using Media Control Function software (Digivox).

Scanning protocol

Imaging was performed on a 1.5-Tesla Siemens Vision scanner. After obtaining high-resolution T1 anatomical scans, two series of 128 gradient-echo images of BOLD signal were acquired (TE = 50 ms, head coil, matrix size: 64×64 ; voxel size: $5 \times 5 \times 5$ mm; 14 slices acquired in the orientation of the Sylvian fissure). A long, 8-s interacquisition interval (TR) was used to ensure low signal contamination by noise artifacts of image acquisition (Belin et al., 1999). Two stimuli were presented during each 8-second TR, and the two trials included in each TR always consisted of the same condition. The following are the temporal parameters of the stimulus presentation times relative to the acquisition. The acquisition lasted 1.4 s, and a 250-ms consonant–vowel (CV) stimulus was presented 0.1 s after the end of the acquisition period. There was then a 3-s silent period during which subjects made their classification response. The second CV was then presented, again followed by 3 s of silence, after which the next image was acquired.

Based on previous research (Belin et al., 1999), we expected the optimal BOLD signal response in the primary auditory cortex (AI) to occur 2 to 3 s after stimulus presentation, and for the response in the secondary auditory cortex to occur approximately 1 s after the response in AI. The hemodynamic response to the acoustic noise of the scanner will likely have faded by the time the next image is acquired. The relative contribution of the first and second trials within each TR to the image will depend on the lag of the BOLD response in the areas of interest and on the subjects' reaction time.

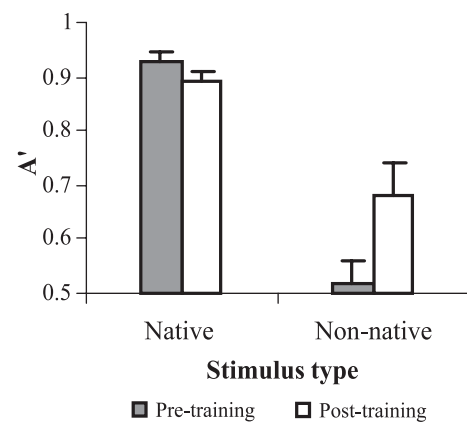


Fig. 2. Pretraining and posttraining identification performance (A').

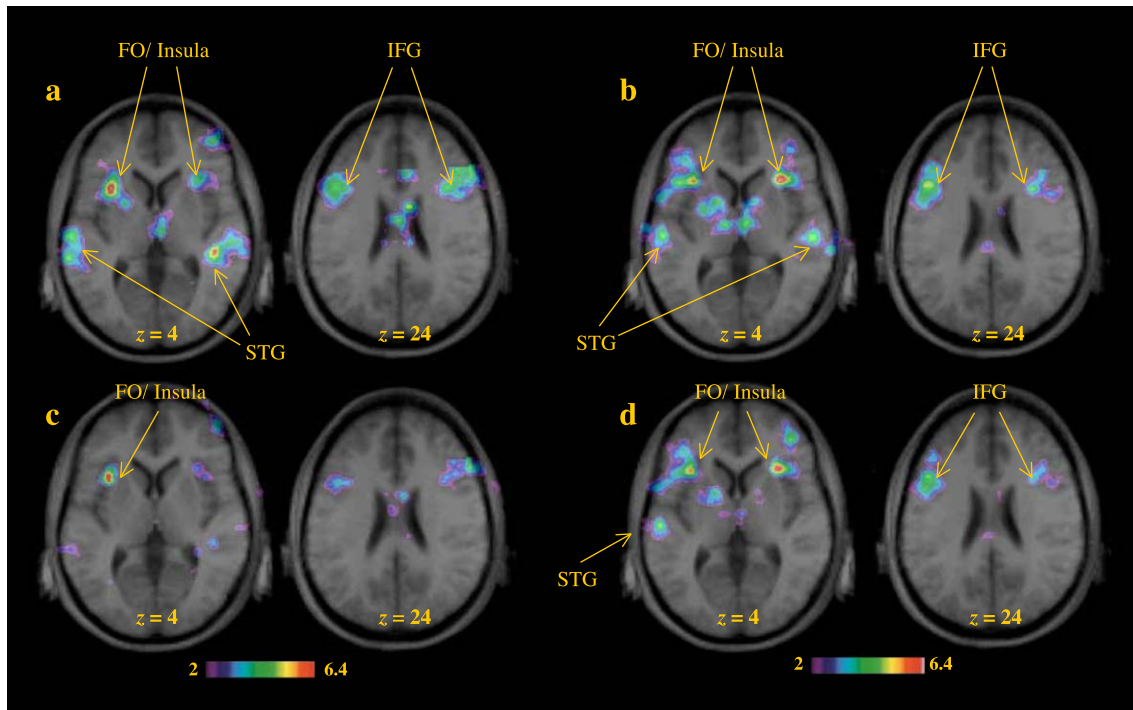


Fig. 3. Statistical parametric maps of fMRI BOLD signal changes (t values, in color scale) superimposed upon group-average anatomical MR image at selected locations for pretraining and posttraining conditions. Each panel shows two horizontal sections taken at corresponding points (indicated by z coordinate in standardized stereotaxic space) for each of the four conditions: (a) pretraining native minus noise, (b) posttraining native minus noise, (c) pretraining nonnative minus noise, and (d) posttraining nonnative minus noise. Abbreviations: FO: frontal operculum; IFG: inferior frontal gyrus; STG: superior temporal gyrus.

Behavioral testing and training procedures

Identification task during pre- and posttraining scans

During each of the two echo-planar imaging (EPI) series, 32 images were acquired during each of the following four conditions: native identification, nonnative identification, noise baseline, and silent baseline. The conditions were distributed such that each of the three conditions was presented during sets of four consecutive acquisitions, followed sometimes by one and sometimes by two silent baseline acquisitions. Conditions were pseudo-randomly ordered. Before the initial scanning session, subjects were briefly familiarized with the stimuli to ensure that they could accurately identify the two native sounds (/da/ and /ta/) and that they could not distinguish the nonnative retroflex /da/ sound from the native dental /da/ sound. They heard each of the dental /da/ exemplar (endpoint sound), the retroflex /da/ exemplar, and the native /ta/ sound 10 times, and were asked to report what they heard. They were also specifically asked if they could hear a difference between the dental and retroflex /da/ sounds.

During the native identification task, subjects were required to identify the /da/ versus /ta/ sounds and to indicate their response by pressing one of two mouse button keys. During the nonnative condition, subjects heard either the dental or retroflex prototypes (i.e., stimulus 1 or 7, respectively). The dental /da/ token (stimulus 1) was identical in the native and non-native conditions. Subjects were not explicitly told about the nonnative condition in the pretraining session. Before training, no subjects could distinguish the dental from the retroflex CVs. For this reason, although they were actively identifying the sounds in both the native and

nonnative conditions (i.e., they were prepared to press the key corresponding to /ta/ if and when it was presented), during the pretraining nonnative condition, subjects in effect pressed the mouse key corresponding to the dental /da/ sounds after each stimulus. After training, they were required to identify the dental versus retroflex sounds (by again pressing mouse button keys). During the noise baseline condition, subjects were to press alternating mouse button keys to the noise bursts.

Phonetic training

Five 1-h sessions of identification training with feedback spread over the course of 2 weeks every 3 days on average were administered between the two scanning sessions. The range across subjects of time between the first and second scans was 2 weeks (1 to 3 weeks). The second always occurred within several days after the last training session.

During the first session, subjects heard endpoint stimuli (sound 1 or 7), and indicated their response by pressing a mouse button. Feedback was presented on the computer screen as to the accuracy of their response. If criterion performance, defined as at least 16/20 correct responses on a block, occurred during three consecutive training blocks, then the task was made more difficult by requiring subjects to identify sound 2 versus 6 (these sounds are a smaller step-size apart, and are acoustically more similar than are the prototypes). The subjects who achieved criterion on this five-step pair were then trained on an even more difficult three-step pair, using stimulus 3 versus 5. This “perceptual fading” training method has been shown to result in behavioral improvement in previous studies (Jamieson and Morosan, 1986; Morosan and

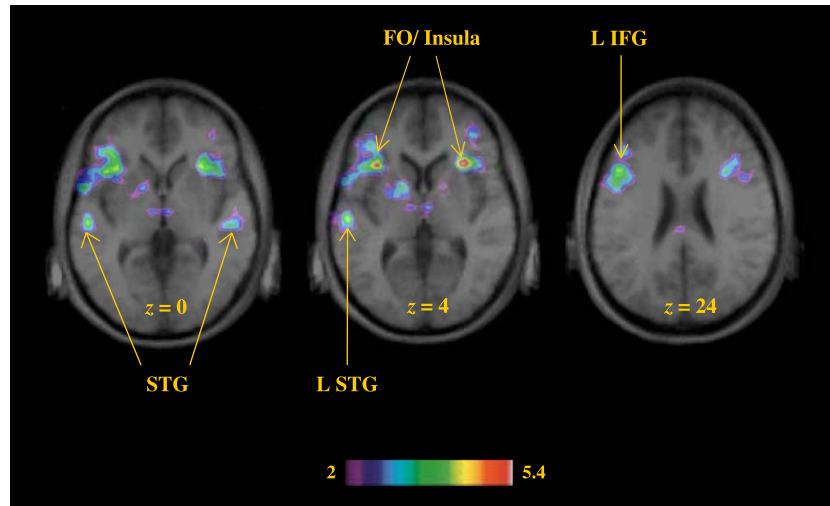


Fig. 4. Result of conjunction analysis illustrating areas of overlap in the posttraining scans between the nonnative and native minus noise subtractions. Abbreviations as in Fig. 3.

Jamieson, 1986). It allows individuals who successfully learn the contrast early during training to then learn additional information regarding more subtle perceptual aspects of the acoustic parameters which critically distinguish the prototypical stimulus pair which they will later be tested with.

Analysis

BOLD signal images were smoothed (6-mm FWHM), corrected for motion artifact and transformed into standard stereotaxic space (Talairach and Tournoux, 1988) using in-house software (Collins et al., 1994). Statistical t maps were obtained in each individual using a voxel-based analysis based on a linear model with correlated errors, with direct specification of the design matrix (Worsley et al., 2001). Output across runs within sessions was combined, and then that of subjects within a population in a hierarchical random effects analysis. Runs were combined with another linear model for the run effects, using Residual Maximum Likelihood (REML) estimation fitted by the Estimation–Maximization (EM) algorithm. The t map images were then registered with a standardized anatomical image, averaged across the subjects. Criteria of $t = 5.02$ in the whole brain volume and of $t = 4.14$ in the predicted areas were calculated (Worsley et al., 1996) based on 374 degrees of freedom, a voxel size of 8 mm^3 , smoothness of 6-mm, and a significance level of $P < 0.05$. The volume of interest for the whole brain was

set as 1000 cc. The threshold for predicted regions was based on searching for the nearest cluster to a prechosen voxel or region of interest.

Results

Behavioral results

During familiarization, subjects could identify the native /da/ and /ta/ sounds 100% of the time. None of the subjects could hear any difference between the dental /da/ and retroflex /da/ sounds, all subjects identified both of these sounds as the dental /da/. The following are the behavioral results of identification performance during scanning. One out of the 10 subjects did not respond to over 50% of the pre- and posttraining identification trials, therefore we excluded this subject's results from all of the analyses. Each subject's performance was converted to an A' score, which is a nonparametric unbiased index of sensitivity that ranges from 0–1.0, with 0.5 corresponding to chance performance (McNichol, 1972). Mean A' scores and standard errors for the native and the nonnative contrasts at pre- and at posttraining are provided in Fig. 2. A (2×2) two-way repeated-measures ANOVA, with time (pre- and posttraining) and stimulus (native and nonnative) being the two independent variables, was per-

Table 1
Native versus noise subtraction

Structure	Pretraining:				BA	Posttraining:			
	Talairach coordinates			t value		Talairach coordinates			t value
	x	y	z			x	y	z	
Left insula–frontal operculum	–34	16	4	6.4	In/44	–34	20	4	5.05
Left inferior frontal gyrus	–48	14	24	4.50	44	–42	26	20	4.51
Right insula–frontal operculum	32	22	10	5.45	In/44	30	20	–2	5.58
Right middle frontal gyrus					46	38	40	14	4.87
Left superior temporal gyrus	–66	–38	4	4.63	22	–56	–22	–2	5.18
Right superior temporal gyrus	46	–32	4	5.94	22	54	–24	0	5.47

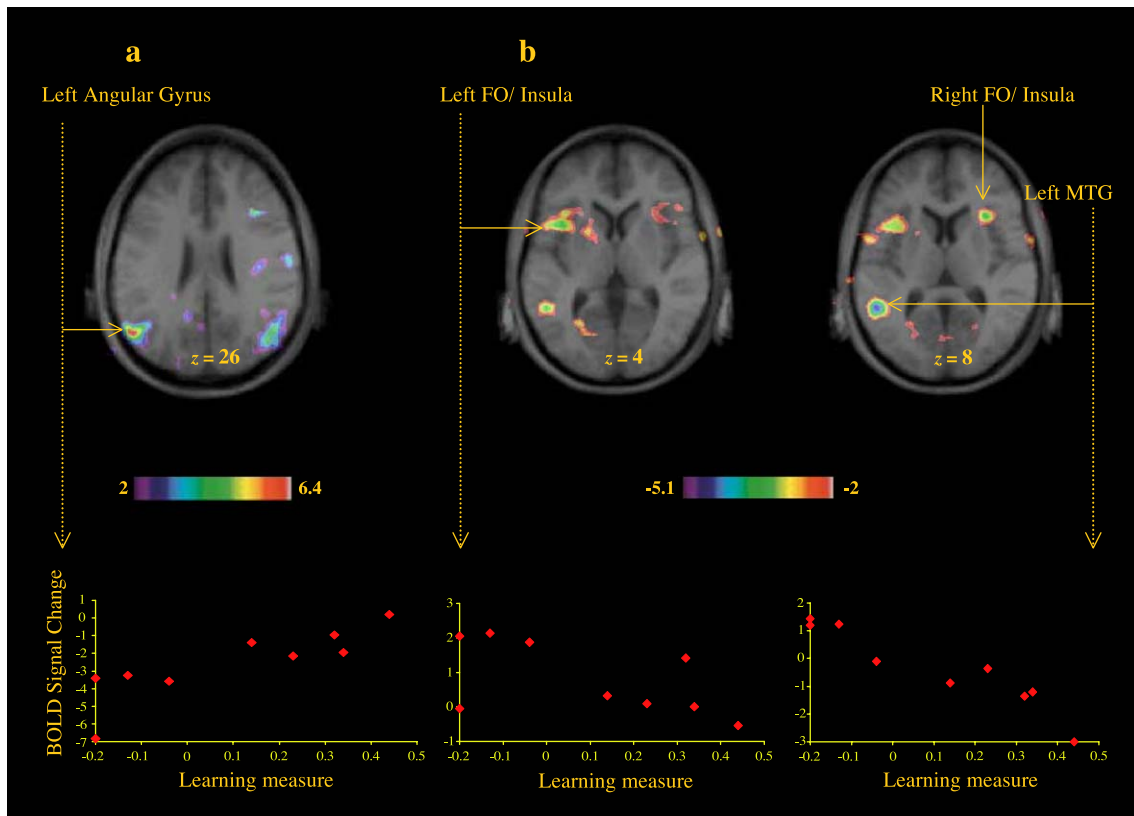


Fig. 5. Correlational analysis between a learning measure (post- minus pretraining A' scores) and the BOLD signal change during the posttraining nonnative minus noise subtraction. Scatter diagrams show the relationship between the BOLD signal change at the peak t value voxel location and the learning measure. (a) Positive correlation between learning measure and activity in angular gyrus; (b) negative relationship between learning measure and activity in bilateral insula–frontal opercular areas and in posterior left middle temporal gyrus (MTG).

formed on the A' scores. Results revealed a significant main effect of stimulus [$F(1,8) = 61.38, P < 0.001$], indicating that overall, performance was better for the native control condition than for the nonnative one, and a significant time by stimulus interaction [$F(1,8) = 13.34, P < 0.01$]. Post hoc Tukey tests on the significant interaction revealed that the posttraining nonnative mean A' measure was significantly higher than the pretraining one [$Q(4,8) = 8.02, P < 0.05$], indicating that subjects' performance improved as a result of training while the native performance remained near ceiling. There was a large range of individual differences with respect to the amount of improvement resulting from training; posttraining A' values for the identifica-

tion of the nonnative contrast ranged from 0.30 to 0.87. Further, a t test for the significance of differences between correlated variances in performance showed that variability in performance is greater in the posttraining nonnative condition as compared to the native condition [$t(7) = 3.88, P < 0.01$].

Imaging Results

Group subtractions

Table 1 provides the stereotaxic coordinates for the pre- and posttraining native minus noise voxel-based subtractions (Fig. 3a). During the pretraining scan, the native minus noise subtraction

Table 2
Nonnative versus noise subtraction

Structure	Pretraining:				BA	Posttraining:			
	Talairach coordinates:			t value		Talairach coordinates:			t value
	x	y	z			x	y	z	
Left insula–frontal operculum	−34	16	4	5.44	In/44	−34	20	4	5.01
Left inferior frontal gyrus					44	−44	16	0	4.54
Right insula–frontal operculum					In/44	30	22	4	5.36
Right middle frontal gyrus					46	40	44	8	5.67
Left superior temporal gyrus					22	−58	−22	2	4.6
Right superior temporal gyrus					22	52	−24	−1	3.68 ^a
Left caudate nucleus						−12	−2	16	5.14

^a Indicates a subthreshold trend.

yielded significant positive activations bilaterally in the insula/frontal opercular (FO) regions and in the superior temporal gyri (STG), and in the left inferior frontal gyrus (IFG). Superior temporal activations, based on inspection of probabilistic maps for Heschl's gyrus (Penhune et al., 1996) and for the planum temporale (Westbury et al., 1999), appeared to be at the junction of these two structures.

The same subtraction after training (Table 1, Fig. 3b) yielded a similar pattern of results with the exception of additional activation in the right frontal gyrus, at a location not symmetrical to the left-sided inferior frontal activation.

Table 2 and Figs. 3d and 3e present the pre- and posttraining results for the nonnative minus noise subtraction. During pretraining, the only significant result was a positive activation in the left insula–frontal operculum. After training, the subtraction yielded positive activations bilaterally in insular–frontal opercular areas, as well as in the left IFG, STG, and caudate nucleus. The right STG activation was only marginally significant but is reported because it falls within 2 mm of the corresponding peak in the native comparison. Additionally, there was a right frontal activation which was once again considerably more anterior than that in the left hemisphere. Conjunction analyses were performed to determine which regions were commonly activated after training in the nonnative minus noise and in the native minus noise subtractions (Table 3, Fig. 4). Results revealed that bilateral frontal operculum–insula and the left superior temporal regions were commonly activated in both of these subtractions. Additionally, there were nonsignificant trends in the left inferior frontal gyrus, in the right superior temporal gyrus, and in the left caudate nucleus.

To directly compare activity associated with identifying the nonnative contrast before versus after training, we subtracted the pretraining nonnative minus noise difference image from the posttraining difference image. In this compound comparison, we found significant activation in the left inferior frontal gyrus and in the left caudate nucleus. Additionally, there were nonsignificant trends in the right inferior frontal gyrus, in the left superior parietal gyrus, and in the left supramarginal gyrus (refer to Table 4). We also performed a compound comparison as a control on the native condition. Comparison of the posttraining native minus noise difference image from the pretraining one did not reveal any significant differences.

Correlation of brain activity with learning

To examine the relationship between brain activity and the degree of learning across individuals, we performed a correlational analysis between posttraining A' minus pretraining A' scores and the BOLD signal in the whole brain volume during the posttraining

Table 3

Conjunction analysis: areas commonly activated by posttraining nonnative minus noise and native minus noise subtractions

Structure	Talairach coordinates:			<i>t</i> value	BA
	<i>x</i>	<i>y</i>	<i>z</i>		
Left insula–frontal operculum	−34	20	4	4.99	In/44
Left inferior frontal gyrus	−49	13	24	3.88 ^a	44
Right insula–frontal operculum	32	22	4	5.36	In/44
Left superior temporal gyrus	−58	−22	2	4.62	22
Right superior temporal gyrus	52	−24	0	3.70 ^a	22
Left caudate nucleus	−18	−3	12	3.60 ^a	

^a Indicates a subthreshold trend.

Table 4

Compound comparison: posttraining nonnative versus noise comparison minus pretraining nonnative versus noise comparison

Structure	Talairach coordinates			<i>t</i> value	BA
	<i>x</i>	<i>y</i>	<i>z</i>		
Left inferior frontal gyrus	−44	26	20	4.32	45
Right inferior frontal gyrus	44	38	8	4.04 ^a	45/46
Left superior parietal gyrus	−26	−74	38	4.19 ^a	7
Left supramarginal gyrus	−44	−36	36	3.71 ^a	40
Left caudate nucleus	−12	4	14	4.69	

^a Indicates a subthreshold trend.

nonnative minus noise subtraction (refer to Table 5). We found positive correlations between this behavioral measure of learning and activity in regions within the left ($r = 0.84$, $P < 0.005$) and right angular gyri (Fig. 5a). Additionally, there were marginally significant negative correlations between the learning measure and activity in bilateral insula–frontal opercular areas, at similar locations to the signal found in the group subtraction results (Fig. 5b). The value of the correlation in the left frontal operculum–insula is $r = -0.55$ ($P = 0.13$); this correlation becomes significant when one outlier is removed ($r = -0.83$, $P < 0.05$). Finally there was a significant inverse relationship between the learning measure and activity in the posterior part of the left middle temporal gyrus ($r = -0.95$, $P < 0.001$) (Fig. 5b). To ensure that the above findings were not due to differences in activation across subjects during the noise baseline condition, we performed a second correlational analysis which showed that the BOLD signal in the posttraining noise minus silence subtraction did not correlate with the rate of learning in any of the regions detected in the original correlational analysis. This suggests that the correlational findings are not due to differences in activation due to the amount of resources dedicated to task-independent conceptual processing during the noise baseline condition.

To test a functional connectivity hypothesis arising from these results (see Discussion), we performed tests of interregional correlations across the areas reported above. Results reveal that when the area of interest is placed in the left frontal operculum–insula at the peak correlation location ($x = -40$, $y = 10$, $z = 4$), there is a significant covariation in activity in the left posterior middle temporal gyrus, at the location: $x = -52$, $y = -48$, $z = 10$ ($t = 4.9$,

Table 5

Correlation between learning measure (post-A' minus pre-A') and posttraining nonnative minus noise subtraction

Structure	Talairach coordinates:			<i>t</i> value	BA
	<i>x</i>	<i>y</i>	<i>z</i>		
<i>Positive correlations</i>					
Left angular gyrus	−54	−66	26	5.47	39
Right angular gyrus	44	−70	34	5.11	39
<i>Negative correlations</i>					
Left insula–frontal operculum	−40	10	4	−3.8 ^a	In/44
Right insula–frontal operculum	30	20	8	−3.76 ^a	In/44
Left poster, middle temporal gyrus	−52	−50	8	−5.13	21

^a Indicates a subthreshold trend.

$P < 0.05$). When the area of interest is placed in the left angular gyrus at the peak correlation location ($x = -54$, $y = -66$, $z = 26$), there is a significant negative covariation in activity in the left posterior middle temporal gyrus, at the location $x = -52$, $y = -46$, $z = 10$ ($t = -5.7$, $P < 0.05$).

Discussion

Behavioral results

The behavioral results followed the expected pattern. They indicate that the training procedure was effective in producing an overall improvement in subjects' identification of the dental–retroflex contrast during the posttraining relative to the pretraining fMRI test sessions, although not all subjects learned to the same extent. This finding is consistent with results of a previous behavioral study (Golestani et al., submitted for publication), in which we showed, using the same paradigm as used here, that training with the dental–retroflex Hindi contrast improved the ability of a larger group ($n = 59$) of nonnative speakers to identify the dental and retroflex sounds, and that those who learned tended to display categorical discrimination functions for the retroflex sounds.

Imaging results

Classifying native speech sounds

As predicted, we found that the classification of native speech sounds recruits regions including the left insula–frontal operculum, the left inferior frontal gyrus, and the left and right superior temporal gyri. We also observed activation in right frontal regions, which was not expected. This result is consistent, however, with those of several previous studies that have shown bilateral frontal activation during the performance of some language tasks (Chee et al., 2001; Fiez et al., 1995).

The bilateral superior temporal activation is consistent with previous findings of bilateral STG involvement in processing both speech and certain types of nonspeech sounds (Binder et al., 1994, 1996; Jäncke et al., 1998, see Hickok and Poeppel, 2000; Zatorre and Binder, 2000, for reviews). These studies have generally shown that secondary auditory regions are bilaterally activated by speech stimuli but not by simple auditory stimuli such as simple tones or noise bursts. Several of these authors have proposed that the primary auditory cortex contributes to the early acoustic processing of all auditory stimuli, whereas the associative areas are involved in higher order processing of sounds. Our finding of greater secondary auditory activation outside of Heschl's gyrus in the phonetic relative to the noise condition is consistent with this explanation.

We found two distinct frontal activations; one in the inferior frontal gyrus (IFG), likely within BA44, and the other in the insula–frontal operculum (FO). Previous functional imaging studies examining phonetic perception have more typically shown activation in regions that converge around the location of our IFG activation (Burton et al., 2000; Démonet et al., 1992, 1994b; Zatorre et al., 1992, 1996). Other studies, however, have shown activation at locations similar to our second insula–FO activation (Chee et al., 2001; Fiez et al., 1995). These two regions likely represent functionally distinct subregions within Broca's speech area (Binder et al., 1997; Buckner et al., 1995; Fiez, 1997; Gold and Buckner, 2002; Zatorre et al., 1996).

According to classical models of speech, the left inferior frontal Broca's area is thought to be involved in the programming and preparation of speech output. There is accumulating evidence, however, that regions in and around the IFG are activated during purely receptive phonetic tasks (Burton et al., 2000; Démonet et al., 1992; Zatorre et al., 1992). Results of functional imaging studies have supported the idea that Broca's area is involved in subvocal rehearsal in verbal working memory (Démonet et al., 1994b; Paulesu et al., 1993). Another prominent explanation for the role of frontal areas derives from the motor theory of speech perception (Lieberman and Mattingly, 1985), which proposes that phonetic perception arises from a system which detects articulatory gestures, and which predicts activation of speech production substrates during certain purely perceptual speech tasks. Consistent with this theory, Zatorre et al. (1992, 1996) proposed that the left posterior temporal regions may be involved in the initial stage of phonetic analysis during passive listening to speech, while the regions including Broca's area are additionally involved when phonetic segments must be extracted and manipulated in relating the phonetic information to articulation. This hypothesis was supported by an fMRI study by Burton et al. (2000), which suggested that the IFG is recruited during phonological processing tasks only when phonetic segmentation or working memory processes are required.

Learning new phonetic contrasts

During the pretraining scan, the nonnative versus noise comparison yielded significant activation only in the left insula–frontal operculum (Table 2 and Fig. 3c). As predicted, after training, the pattern of activation in this comparison was more similar to that in the native versus noise comparison (Table 2 and Fig. 3d), except for the additional activation of the left caudate nucleus in the nonnative relative to the native posttraining condition.

Evidence that the caudate nucleus, and the basal ganglia more generally, may be involved in speech comes from neuropsychological work showing pathology of the putamen and caudate nuclei in aphasia (Lieberman et al., 1992; Pickett et al., 1998; Ullman et al., 1997). Damage to these regions can lead to oral and verbal dyspraxias (Aglioti et al., 1996; Blumstein et al., 1987; Speedie et al., 1993). Additionally, there is evidence from brain imaging studies of basal ganglia involvement, in particular of the putamen, in the articulation and motor control of speech (Klein et al., 1994; Wise et al., 1999). Our finding of left caudate nucleus activation in the posttraining nonnative condition may thus be related to greater subvocal articulatory demands in the perception of this newly learned contrast, even though no speech output was required.

The similarity between the posttraining native and nonnative activation patterns was confirmed by the results of a conjunction analysis comparing the two posttraining conditions (Table 3 and Fig. 4). Significant overlap across the training conditions was observed bilaterally in the frontal opercula and in the left superior temporal gyrus, and a nonsignificant trend was found in the right superior temporal gyrus. To interpret these findings, it is important to recognize that the nonnative identification task is performed differently after relative to before learning, since an inherent consequence of learning is a change in behavior. For example, although subjects were performing an identification task in both conditions, some subjects could in one case (posttraining) hear the difference between the sounds whereas they could not in the other case (pretraining). This makes it likely the task was performed

more ‘automatically’ in the ‘nonnative’ blocks before relative to after training. There are methodological and conceptual problems underlying any study of plasticity, whether it be due to development or to learning (see Poldrack, 2000, for a theoretical discussion and review of these issues). Performance confounds are by definition condition-specific; for example, changes in reaction time or in task difficulty can alter areas activated and/or hemodynamics of activation in particular regions.

Related to the above issues, given that our participants could not hear the difference between the dental and retroflex sounds before training, they effectively heard the dental sound many more times during the pre- relative to in the posttraining condition. It is therefore possible that there were greater adaptation effects in auditory cortices, in the pre- relative to the posttraining nonnative condition, which may in part explain the lack of activation in the superior temporal gyri before training in the nonnative condition (Grill-Spector and Malach, 2001; Pfeiderer et al., 2002). Despite these limitations with the interpretation of the pretraining nonnative condition, we believe that the post- minus pretraining results at least partly reflect a learning process because some of the same patterns of activation are seen in the correlational analysis, which are based solely on the posttraining data (see below).

These findings have implications for neurolinguistic models of second language learning. As mentioned in the Introduction, the ability to distinguish certain nonnative speech sounds from native ones is diminished during development (Best et al., 1988; Werker and Tees, 1983, 1984a; Werker et al., 1981). Our results suggest that learning a new speech sound in adulthood results largely in the recruitment of the same neural substrates as those involved during the classification of native speech sounds. This interpretation is consistent with results of neuroimaging studies of language function in bilinguals showing that at the single word level, brain regions subserving the native language (L1) and the second language (L2) appear to overlap (Chee et al., 1999; Illes et al., 1999; Klein et al., 1994, 1995). Our results are based on the learning of a single, synthetic, nonnative consonant–vowel (CV) syllable. We predict that we would have found similar results had several nonnative CVs been trained, but these results may or may not generalize to naturalistic stimuli, nor to the lexical and semantic aspects of learning a new language. It also remains to be studied how factors such as proficiency, age of acquisition and similarity of phonetic space between native and nonnative languages would influence changes in brain activity resulting from learning.

In addition, short-term learning effects such as the ones studied here may not entirely predict brain activity associated with expert performance. This is an important caveat given that posttraining performance in our subjects never approached native levels (Fig. 2). It may be the case, for example, that given that the amount of training that our subjects received was limited in duration and that it involved only one nonnative consonant–vowel syllable, the imaging results may reflect neural activity associated with the native speech system trying to accommodate a new set of sounds, rather than a novel pattern of neural activity that would evolve as a result of training specifically designed to subserve nonnative phonetic perception.

Change in brain function associated with learning new speech sounds

To confirm the effects seen in the subtraction analyses presented above, we performed a compound analysis in which we

subtracted the pretraining nonnative minus noise difference image from the posttraining one. This analysis yielded significant activity in the left inferior frontal gyrus and in the left caudate nucleus. The left inferior frontal activation was at a similar location to that found in the posttraining native-minus-noise group subtraction reported above (Table 1), and suggests that classifying newly learned speech sounds recruits this region, as does the classification of native speech sounds. This analysis does not, however, support the group average nonnative-minus-noise finding described above suggesting that the second, more inferior insula–frontal opercular (FO) region is also recruited when newly learned speech sounds are classified. As already reported, we found large individual differences in the behavioral performance of our subjects; only approximately half of our subjects improved in their ability to classify the sounds after training. It is possible that activity in certain brain regions such as the insula–FO increased for some subjects and decreased for others as a function of success in learning, and that such learning-related differences cancelled each other out such that they are not detectable in this compound comparison. This interpretation was supported in the correlational analysis to be described below. In addition, the behavioral data support the hypothesis of greater individual variation in the identification of the nonnative relative to the native phonetic contrast. We found that after training, the variability in performance is indeed greater in the nonnative relative to native conditions.

Correlations of brain activity with learning

We found positive correlations between a behavioral learning measure and activity in the left and right angular gyri (Fig. 5a); scatter plots reveal that in better learners, there is less deactivation of these regions (relative to the noise baseline) than in worse learners. There were also concurrent marginally significant negative correlations with activity in insula–frontal opercular regions bilaterally (Fig. 5b), as well as significant negative correlation with activity in the posterior left middle temporal gyrus; as seen in the scatterplot, better learners deactivate this latter region more than worse learners do. A number of studies have shown involvement of the left middle temporal region during the performance of language-related tasks involving lexico–semantic processing (Binder et al., 1996, 1997, 2000; Démonet et al., 1992, 1994a). Our results suggest that activity in this speech-related region is also modulated by the degree of phonetic learning such that people who are better at learning a novel phonetic contrast suppress this region more than do ones who are poorer at it.

Angular gyrus lesions are well known to produce language deficits, and many studies have demonstrated activation associated with phonological processing in the left temporoparietal regions outside the superior temporal gyrus, including in the angular, supramarginal, and middle and inferior temporal gyri (Binder et al., 1996, 1997; Démonet et al., 1994a; Paulesu et al., 1993; Petersen et al., 1988; Zatorre et al., 1992, 1996). The strong positive correlation between learning and activation in the left angular gyrus supports the idea that activity in this region is modulated by learning, such that poorer learners deactivate this region more than faster learners do. To help interpret this result, we tested whether there was decreased BOLD signal near the angular gyrus in the posttraining nonnative minus noise subtraction, and the result showed that this was the case ($t = -4.17$, $P < 0.05$ at $x = -46$, $y = -62$, $z = 22$). In contrast, such a deactivation was not found in the posttraining native minus noise subtraction for the native speech contrast.

The finding of functional differences in the parietal lobes bilaterally between faster versus slower phonetic learners parallel those of a previous morphological study in which we examined the relationship between brain structure and individual differences in phonetic learning (Golestani et al., 2002). We acquired anatomical magnetic resonance images of 59 healthy right-handed English-speaking individuals (not part of the present sample), and later trained these same subjects to learn the Hindi dental–retroflex nonnative contrast. We examined relationships between brain morphology and learning rate using voxel-based morphometry, and found that in faster learners there was more white than gray matter in the parietal lobes, especially in the left hemisphere. This finding, as well as those of the present study, suggests that bilateral functional and structural differences in the parietal lobes can in part predict phonetic learning in normal individuals. The anatomical correlates, together with the present findings, identify aspects of the functional and structural circuitry predictive of phonetic learning. It appears that interactions occurring within this system are important determinants of success in learning novel mappings between sounds and their internal representations; however, the precise mechanisms underlying possible relationships between morphological and functional differences across individuals in relation to phonetic learning abilities are apparently quite complex and remain to be fully understood.

We had expected a positive correlation between learning and insula–frontal opercular activation because we had predicted that more successful task performance would place greater processing demands on speech regions. In contrast, we found an inverse relationship with bilateral activity in essentially the same insula–frontal opercular regions as were detected in the group subtraction results. This finding suggests that although all subjects recruit these areas for performing the task, good learners engage frontal speech regions to a lesser extent than poor learners do. Individual differences in activity in this region likely resulted in a training-related group average signal that was too weak to be detected in the compound comparison analysis reported above. These results are based on correlations, and do not inform us about the direction of causality, if any, between differences across subjects in brain activation and in performance. We will, however, speculate about the mechanism of this function–behavior relationship, and suggest a number of possible explanations for this finding. The first is that in learners, processing in these frontal regions may be more efficient and/or may require less processing time than in non-learners. Several mechanisms may underlie this difference in functional anatomy. It may be that poor learners do not succeed in developing a stable representation of the new nonnative sound in their long-term memory store, and that this results in a relatively greater involvement of articulatory strategies while subjects attempt to classify the sounds. Additionally, poor learners may engage subvocal rehearsal strategies in phonological verbal working memory relatively more than good learners. In other words, maybe they rely more on the strategy of comparing each sound to the previous one as they attempt to classify the sounds, and are less able than the good learners to match each sound to long-term memory “templates” for the native versus nonnative sounds. The latter explanation is consistent with reports during training by some subjects that they did compare each sound to the previous one, and that their performance tended to deteriorate every time that they received negative feedback on an item.

The hypothesis of more efficient processing in frontal areas in learners is consistent with the results of several studies which have

shown that prefrontal regions are relatively more active in tasks requiring “top-down” processing (Frith et al., 1991; Raichle et al., 1994). These studies have also shown that when processing becomes more “bottom-up”, or automated, there is a decrease in such frontal activity accompanied by an increase in activity in more posterior regions of the brain. In addition, Chee et al. (2001) showed greater activity bilaterally in the insula–frontal operculum during a semantic task in the less proficient of two known languages in bilinguals, and suggested that processing words in the less familiar language requires greater cognitive effort, that it is subserved by less well-tuned neural representations, and that it therefore requires greater neuronal activity than processing words in the more familiar language. We thus suggest a second interpretation, not exclusive with the first, that poor learners continue to engage relatively more “top-down,” effortful cognitive processes, resulting in stronger activation in regions including frontal speech areas. In contrast, in good learners, the task is likely performed in a more “bottom-up” and automated manner after training. According to the work cited above, it might be expected that more automated processing would be accompanied by less activity in frontal regions, as well as by more activity in posterior regions. Results in the frontal regions of the brain fit with this prediction, however, rather than observing more activity in posterior regions in better learners, we observe that better learners appear to deactivate posterior regions less than poorer learners. This result can be interpreted as an extension of previous work: maybe when tasks are more effortful and novel, activity in posterior regions is not only less strong but maybe it is even more suppressed relative to when a task is more automated or bottom-up.

It is further possible that there are differences between good and poor learners in the degree of functional connectivity between the insula–FO and parietal regions, and that such differences may be related to underlying differences in white matter morphology. It has been suggested, based on functional imaging (Paulesu et al., 1996) and on microstructural data (Klingberg et al., 2000) that normal functional connectivity between the left inferior parietal regions and Broca’s area is disrupted in dyslexic individuals (Paulesu et al., 1996). Functional connectivity may or may not indicate direct anatomical connectivity between these two regions, since activity in these regions may be modulated by input from a third part of the brain. In the present study, tests of functional connectivity revealed that activity in the left frontal operculum–insula covaries positively with that in the posterior left middle temporal region, and that activity in the left angular region covaries negatively with that in this same posterior temporal area. These results support the interpretation of differences in functional connectivity between individuals that learn new speech sounds more easily or more slowly. Further, it appears that the nature of such differences is more complex than might have initially been predicted. For example, it appears that both functional activation and functional suppression are involved. It would be interesting to test the hypothesis that learning to perceive a new speech sound involves inhibiting prelearned representations of native sounds, and that such an inhibition arises from the functional suppression of one region by another. Alternatively, differences across individuals in the degree of functional suppression of more posterior by more anterior regions might arise from more general cognitive mechanisms such as differences in top-down versus bottom-up processing, and these in turn may be related to differences in working memory or other executive or strategic factors across individuals. Further studies can be designed to address these possibilities.

The group subtraction analysis as well as the correlational analyses yielded significant peaks bilaterally in very similar insula–frontal opercular areas. This suggests that activity in this area is both (a) recruited in the classification task, and (b) modulated by learning. In contrast, activations in the left inferior frontal gyrus (BA 44) and superior temporal regions were only detected in the group subtraction analysis, suggesting that these areas are recruited by performance of the identification task but that they are not modulated by the degree of success in learning. These findings also support the idea presented above that within the frontal speech area, the IFG and the insula/FO areas subserve different aspects of phonetic perception.

Summary and conclusions

We demonstrate overall behavioral improvement, together with large individual differences in performance, in the ability of monolingual English speakers to identify the Hindi dental–retroflex contrast after 5 h of phonetic training. Functional imaging results during the classification of a native control phonetic contrast revealed the involvement of speech-related brain regions including the superior temporal and insula–frontal opercular regions bilaterally. Activity associated with the post-training classification of the newly learned phonetic contrast revealed a pattern of activation very similar to that found during the classification of the native phonetic contrast. This finding has important implications for neuro-linguistic theories of second language learning, and suggests that only 5 h of phonetic training in adulthood with a nonnative phonetic contrast, the perception of which is likely to have been diminished as early as at 6 months of age, is sufficient to recruit the same substrates underlying the perception of speech sounds to which one has been exposed since birth.

In examining differences in brain function across subjects related to differences in the degree of success in learning the new speech sound, we found that in more successful learners, activity in the left middle temporal gyrus is more suppressed relative to less successful learners. Poorer learners, on the other hand, appear to suppress activity in the left angular gyrus more than better learners, and further, they show relatively more activity in insula–frontal opercular regions. These latter result suggests that frontal speech regions are relatively less active in “good learners” relative to in “poor learners”, and that processing in these regions is more efficient and/or may require less processing time when successful phonetic learning has taken place. This in turn suggests that in successful learners, there is either less need for or more automatized processing related to subvocal rehearsal strategies in verbal working memory and/or related to mapping the perceived speech sounds onto an articulatory template. More generally, the concurrent modulation of activity in posterior and anterior speech regions as a function of learning may suggest that the functional connectivity between posterior and anterior speech regions is modulated by learning new speech sounds.

Acknowledgments

We thank Pierre Ahad for help in creating the synthetic stimuli and Rhonda Amsel for statistical advice. We also thank Michael Petrides, Valentina Petre, Pascal Belin, Keith Worsley, Marc Bouffard, and Peter Neelin for technical assistance and consultation,

and Bruce Pike for access to the MNI Brain Imaging Centre facilities.

Funding was provided by the Canadian Institutes of Health Research (Operating grants 11541 and 14995) and by the McDonnell-Pew Cognitive Neuroscience Program.

References

- Aglioti, S., Beltramello, A., Girardi, F., Fabbro, F., 1996. Neurolinguistic and follow-up study of an unusual pattern of recovery from bilingual subcortical aphasia. *Brain* 119, 1551–1564.
- Alho, K., Winkler, I., Escera, C., Huottilainen, M., Virtanen, J., Jääskeläinen, I.P., Pekkonen, E., Ilmoniemi, R.J., 1998. Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology* 35, 211–224.
- Belin, P., Zatorre, R.J., Hoge, R., Pike, B., Evans, A.C., 1999. Event-related fMRI of the auditory cortex. *NeuroImage* 10, 417–429.
- Benson, D.F., 1967. Fluency in aphasia. *Cortex* 3, 373–394.
- Benson, D.F., Sheremata, W.A., Bouchard, R., Segarra, J.M., Price, D., Geschwind, N., 1973. Conduction aphasia: a clinicopathological study. *Arch. Neurol.* 28, 339–346.
- Best, C.T., McRoberts, G.W., Sithole, N.M., 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *J. Exp. Psychol. Hum. Percept. Perform.* 14 (3), 345–360.
- Binder, J.R., Rao, S.M., Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandettini, P.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., Haughton, V.M., Hyde, J.S., 1994. Functional magnetic resonance imaging of human auditory cortex. *Ann. Neurol.* 35 (6), 662–672.
- Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M., Cox, R.W., 1996. Function of the left planum temporale in auditory and linguistic processing. *Brain* 119, 1239–1247.
- Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., Prieto, T., 1997. Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17 (1), 353–362.
- Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and non-speech sounds. *Cereb. Cortex* 10, 512–528.
- Blumstein, S.E., Baker, E., Goodglass, H., 1977. Phonological factors in auditory comprehension aphasia. *Neuropsychologia* 15, 19–30.
- Blumstein, S.E., Alexander, M.P., Ryalls, J.H., Katz, W., Dworetzky, B., 1987. On the nature of the foreign accent syndrome: a case study. *Brain Lang.* 31, 215–244.
- Buckner, R.L., Raichle, M.E., Petersen, S.E., 1995. Dissociation of human prefrontal cortical areas across different speech production tasks and gender groups. *J. Neurophysiol.* 74, 2163–2173.
- Burnham, D.K., 1986. Developmental loss of speech perception: exposure to and experience with a first language. *Appl. Psycholinguist.* 7, 207–240.
- Burton, M.W., Small, S.L., Blumstein, S.E., 2000. The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12 (4), 679–690.
- Chee, M.W.L., Tan, E.W.L., Thiel, T., 1999. Mandarin and English single word processing studied with functional magnetic resonance imaging. *J. Neurosci.* 19 (8), 3050–3056.
- Chee, M.W.L., Hon, N., Lee, H.L., Soon, C.S., 2001. Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *NeuroImage* 13, 1155–1163.
- Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205.
- Csépe, V., 1995. On the origin and development of the mismatch negativity. *Ear Hear.* 16, 91–104.
- Démonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., Frackowiak, R., 1992. The anatomy of phono-

- logical and semantic processing in normal subjects. *Brain* 115, 1753–1768.
- Démonet, J.F., Price, C., Wise, R., Frackowiak, R.S.J., 1994a. Differential activations of right and left posterior sylvian regions by semantic and phonological tasks: a positron emission tomography study in normal human subjects. *Neurosci. Lett.* 182, 25–28.
- Démonet, J.F., Price, C., Wise, R., Frackowiak, R.S.J., 1994b. A PET study of cognitive strategies in normal subjects during language tasks: influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain* 117, 671–682.
- Fiez, J.A., 1997. Phonology, semantics, and the role of the left inferior prefrontal cortex. *Hum. Brain Mapp.* 5, 79–83.
- Fiez, J.A., Raichle, M.E., Miezin, F.M., Petersen, S.E., 1995. PET studies of auditory and phonological processing: effects of stimulus characteristics and task demands. *J. Cogn. Neurosci.* 7 (3), 357–375.
- Flege, J.E., 1984. The effect of linguistic experience on Arabs' perception of the English /s/ vs /z/ contrast. *Folia Linguist.* 18, 117–138.
- Frith, C.D., Friston, K.J., Liddle, P.F., Frackowiak, R.S.J., 1991. A PET study of word finding. *Neuropsychologia* 29 (12), 1137–1148.
- Geschwind, N., 1970. The organization of language and the brain. *Science* 170, 940–944.
- Geschwind, N., 1971. Aphasia. *New Engl. J. Med.* 284, 654–656.
- Gold, B.T., Buckner, R.L., 2002. Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron* 35, 803–812.
- Golestani, N., Paus, T., Zatorre, R.J., 2002. Anatomical correlates of learning novel speech sounds. *Neuron* 35, 997–1010.
- Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol.* 107, 293–321.
- Henke, W.L., 1990. MITSYN: a synergistic family of high-level languages for time signal processing, Version 8.1 Author, Belmont, MA.
- Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138.
- Illes, J., Francis, W.S., Desmond, J.E., Gabrieli, J.D.E., Glover, G.H., Poldrack, R., Lee, C.J., Wagner, A.D., 1999. Convergent cortical representation of semantic processing in bilinguals. *Brain Lang.* 70, 347–363.
- Jamieson, D.G., Morosan, D.E., 1986. Training nonnative speech contrasts in adults: acquisition of the English /s/-/?/ contrast by francophones. *Percept. Psychophys.* 40 (4), 205–215.
- Jäncke, L., Shah, N.J., Posse, S., Grosse-Ryken, M., Müller-Gärtner, H.W., 1998. Intensity coding of auditory stimuli: an fMRI study. *Neuropsychologia* 36, 875–883.
- Jusczyk, P.W., 1995. Language acquisition: speech sounds and the beginning of phonology. In: Miller, J.L., Eimas, P.D. (Eds.), *Speech, Language, and Communication*. Academic Press, San Diego, pp. 263–301.
- Klein, D., Zatorre, R.J., Milner, B., Meyer, E., Evans, A.C., 1994. Left putamen activation when speaking a second language: evidence from PET. *NeuroReport* 5 (17), 2295–2297.
- Klein, D., Milner, B., Zatorre, R.J., Meyer, E., Evans, A.C., 1995. The neural substrates underlying word generation: a bilingual functional-imaging study. *Proc. Natl. Acad. Sci. U. S. A.* 92, 2899–2903.
- Klingberg, T., Hedehus, M., Temple, E., Salz, T., Gabrieli, J.D., Moseley, M.E., Poldrack, R.A., 2000. Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron* 25 (2), 493–500.
- Kraus, N., Cheour, M., 2000. Speech sound representation in the brain. *Audiol. Neuro-Otol.* 5, 140–150.
- Kraus, N., Disterhoft, J., 1982. Response plasticity of single neurons in rabbit auditory association cortex during tone-signalled learning. *Brain Res.* 246, 205–215.
- Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., Nicol, T., 1995. Central auditory system plasticity associated with speech discrimination training. *J. Cogn. Neurosci.* 7, 25–32.
- Kuhl, P.K., 2000. A new view of language acquisition. *Proc. Natl. Acad. Sci.* 97 (22), 11850–11857.
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by six months of age. *Science* 255, 606–608.
- Lieberman, A.M., 1957. Some results of research on speech perception. *J. Acoust. Soc. Am.* 29, 117–123.
- Lieberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. *Cognition* 21, 1–36.
- Lieberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C., 1957. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368.
- Lieberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. *Psychol. Rev.* 74, 431–461.
- Lieberman, P., Kako, E., Friedman, J., Tajchman, G., Feldman, L.S., Jimenez, E.B., 1992. Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain Lang.* 43, 169–189.
- Mazoyer, B.M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., Mehler, J., 1993. The cortical representation of speech. *J. Cogn. Neurosci.* 5 (4), 467–479.
- McNichol, D., 1972. *A Primer of Signal Detection Theory*. Allen & Unwin, London.
- Morosan, D.E., Jamieson, D.G., 1986. Efficient training of nonnative speech discrimination with perceptual fading. *Proceedings of the 12th International Congress on Acoustics*. Canadian Acoustical Association.
- Mummery, C.J., Ashburner, J., Scott, S.K., Wise, R.J., 1999. Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* 106, 449–457.
- Näätänen, R., Gaillard, A., Mäntysali, S., 1978. Early selective attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329.
- Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., Paavilainen, P., 1993. Development of a memory trace for a complex sound in the human brain. *NeuroReport* 4, 503–506.
- Pantev, C., Wollbrink, A., Roberts, L.E., Engelien, A., Lütkenhöner, B., 1999. Short-term plasticity of the human auditory cortex. *Brain Res.* 842, 192–199.
- Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. *Nature* 362, 342–345.
- Paulesu, E., Frith, U., Snowling, M., Gallagher, M., Morton, J., Frackowiak, R.S.J., Frith, C.D., 1996. Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain* 119, 143–157.
- Penhune, V.B., Zatorre, R.J., MacDonald, J.D., Evans, A.C., 1996. Inter-hemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurements from magnetic resonance scans. *Cereb. Cortex* 6, 661–672.
- Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E., 1988. Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature* 331, 585–589.
- Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E., 1989. Positron emission tomography studies of the processing of single words. *J. Cogn. Neurosci.* 1, 153–170.
- Pfleiderer, B., Ostermann, J., Michael, N., Heindel, W., 2002. Visualization of auditory habituation by fMRI. *NeuroImage* 17, 1705–1710.
- Pickett, E.R., Kuniholm, E., Protopapas, A., Friedman, J., Lieberman, P., 1998. Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: a case study. *Neuropsychologia* 36, 173–188.
- Pisoni, D.B., Aslin, R.N., Perey, A.J., Hennessy, B.L., 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 297–314.
- Poldrack, R.A., 2000. Imaging brain plasticity: conceptual and methodological issues—A theoretical review. *NeuroImage* 12, 1–13.
- Polka, L., 1991. Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. *J. Acoust. Soc. Am.* 89 (6), 2961–2977.
- Polka, L., Werker, J.F., 1994. Developmental changes in the perception of nonnative vowel contrasts. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 421–435.
- Priutt, J.S., Strange, W., Polka, L., Aguilar, M.C., 1990. Effects of category

- knowledge and syllable truncation during auditory training on Americans' discrimination of Hindi retroflex–dental contrasts. *J. Acoust. Soc. Am.* 87 (Suppl. 1), S72.
- Raichle, M.E., Fiez, J.A., Videen, T.O., MacLeod, A.K., Pardo, J.V., Fox, P.T., Petersen, S.E., 1994. Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex* 4, 8–26.
- Recanzone, G., Schreiner, C., Merzenich, M., 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13, 87–104.
- Rumsey, J.M., Andreason, P., Zemetkin, A.J., Aquino, T., King, C., Hamburger, S.D., Pikus, A., Rapoport, J.L., Cohen, R.M., 1992. Failure to activate the left temporoparietal cortex in dyslexia. *Arch. Neurol.* 49, 527–534.
- Speedie, L.J., Wertman, E., Ta'ir, J., Heilman, K.M., 1993. Disruption of automatic speech following a right basal ganglia lesion. *Neurology* 43, 1768–1774.
- Stevens, K.N., Blumstein, S.E., 1975. Quantal aspects of consonant production and perception: a study of retroflex stop consonants. *J. Phon.* 3, 215–233.
- Strange, W., Dittman, S., 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Percept. Psychophys.* 36 (2), 131–145.
- Strange, W., Polka, L., Aguilar, M.C., 1989. Effects of auditory and phonetic training on Americans' discrimination of Hindi retroflex–dental contrasts. *J. Acoust. Soc. Am.* 86 (Suppl. 1), S101.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme Medical, New York.
- Tallal, P., Newcombe, F., 1978. Impairment of auditory perception and language comprehension in dysphasia. *Brain Lang.* 5, 13–24.
- Tees, R.C., Werker, J.F., 1984. Perceptual flexibility: maintenance or recovery of the ability to discriminate nonnative speech sounds. *Can. J. Psychol.* 38 (4), 579–590.
- Tervaniemi, M., Medvedev, S.V., Alho, K., Pakhomov, S.V., Roudas, M.S., van Zuijen, T.L., Näätänen, R., 2000. Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10, 74–79.
- Tremblay, K., Kraus, N., Carrell, T.D., McGee, T., 1997. Central auditory system plasticity: generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.* 102, 3762–3773.
- Tremblay, K., Kraus, N., McGee, T., 1998. The time course of auditory perceptual learning: neurophysiological changes during speech–sound training. *NeuroReport* 9 (16), 3557–3560.
- Ullman, M.T., Corkin, S., Coppola, M., Hickok, G., Crowdon, J.H., Koro-shetz, W.J., Pinker, S., 1997. A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *J. Cogn. Neurosci.* 9, 266–276.
- Werker, J.F., Lalonde, C.E., 1988. Cross-language speech perception: initial capabilities and developmental change. *Dev. Psychobiol.* 24 (5), 672–683.
- Werker, J.F., Logan, J.S., 1985. Cross-language evidence of three factors in speech perception. *Percept. Psychophys.* 37 (1), 35–44.
- Werker, J.F., Tees, R.C., 1983. Developmental changes across childhood in the perception of nonnative speech sounds. *Can. J. Psychol.* 37, 278–286.
- Werker, J.F., Tees, R.C., 1984a. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63.
- Werker, J.F., Tees, R.C., 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* 75, 1865–1878.
- Werker, J.F., Gilbert, J.H.V., Humphrey, K., Tees, R., 1981. Developmental aspects of cross-language speech perception. *Child Dev.* 52, 349–355.
- Westbury, C.F., Zatorre, R.J., Evans, A.C., 1999. Quantifying variability in the planum temporale: a probability map. *Cereb. Cortex* 9 (4), 392–405.
- Wise, R.J., Greene, J., Buchel, C., Scott, S.K., 1999. Brain regions involved in articulation. *Lancet* 353, 1057–1061.
- Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* 4, 58–73.
- Worsley, K.J., Liao, C., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2001. A general statistical analysis for fMRI data. *NeuroImage* 15 (1), 1–15.
- Zatorre, R.J., Binder, J.R., 2000. Functional and structural imaging of the human auditory system. In: Toga, A.W., Mazziota, J.C. (Eds.), *Brain Mapping: The Systems*. Academic Press, San Diego, pp. 365–402.
- Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846–849.
- Zatorre, R.J., Meyer, E., Gjedde, A., Evans, A.C., 1996. PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb. Cortex* 6, 21–30.