◆ Human Brain Mapping 28:995–1006 (2007) ◆

Neural Characteristics of Successful and Less Successful Speech and Word Learning in Adults

Patrick C.M. Wong,[1]* Tyler K. Perrachione,[2] and Todd B. Parrish[3]

[1]*The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University Institute for Neuroscience, Northwestern University, Evanston, Illinois*
[2]*Department of Linguistics and Cognitive Science Program, Northwestern University, Evanston, Illinois*
[3]*Department of Radiology and Department of Biomedical Engineering, Northwestern University, Chicago, Illinois*

Abstract: A remarkable characteristic of the human nervous system is its ability to learn to integrate novel (foreign) complex sounds into words. However, the neural changes involved in how adults learn to integrate novel sounds into words and the associated individual differences are largely unknown. Unlike English, most languages of the world use pitch patterns to mark individual word meaning. We report a study assessing the neural correlates of learning to use these pitch patterns in words by English-speaking adults who had no previous exposure to such usage. Before and after training, subjects discriminated pitch patterns of the words they learned while blood oxygenation levels were measured using fMRI. Subjects who mastered the learning program showed increased activation in the left posterior superior temporal region after training, while subjects who plateaued at lower levels showed increased activation in the right superior temporal region and right inferior frontal gyrus, which are associated with nonlinguistic pitch processing, and prefrontal and medial frontal areas, which are associated with increased working memory and attentional efforts. Furthermore, we found brain activation differences even before training between the two subject groups, including the superior temporal region. These results demonstrate an association between range of neural changes and degrees of language learning, specifically implicating the physiologic contribution of the left dorsal auditory cortex in learning success. *Hum Brain Mapp 28:995–1006, 2007.* © 2006 Wiley-Liss, Inc.

Key words: learning; brain plasticity; speech perception; auditory perception; language processing; auditory cortex

INTRODUCTION

The neural mechanisms subserving sound-to-word learning in adulthood after years of reduced experience with specific (foreign) speech sounds are largely unknown. Furthermore, although behavioral studies have reported individual differences in second language learning in adulthood, including the extremely low rate of native-like competency (e.g., Bley-Vroman, 1989; Selinker, 1972), few studies have investigated the neural mechanisms associated with differences in learning attainment. In the current study, we investigate neural changes associated with sound-to-word learning in adults. In

Contract grant sponsor: Northwestern University; Contract grant sponsor: National Institutes of Health; Contract grant numbers: HD051827 and DC007468.

*Correspondence to: Patrick C.M. Wong, Ph.D., Dept. of Communication Sciences and Disorders, Northwestern University, 2240 Campus Dr., Evanston, IL 60208. E-mail: pwong@northwestern.edu

Received for publication 20 April 2006; Revised 19 July 2006; Accepted 20 July 2006

DOI: 10.1002/hbm.20330

Published online 28 November 2006 in Wiley InterScience (www.interscience.wiley.com).

addition, we examine neural characteristics of successful and less successful learning before and after such learning.

Neural Bases of Speech and Word Learning in Adults

Several recent studies found learning-related neural changes associated with the learning of foreign speech sounds and words. Generally, these studies were either concerned with the learning of foreign sounds without considering their contribution to larger linguistic contexts, such words (Callan et al., 2003; Golestani and Zatorre, 2004), or with word learning without considering the contribution of specific phonetic features (McLaughlin et al., 2004; Raboyeau et al., 2004). For example, Golestani and Zatorre (2004) found that English subjects learning to identify the Hindi dental-retroflex contrast without actually using those consonants in words resulted in increased activation of the left superior temporal gyrus, insula-frontal operculum, and inferior frontal gyrus. In terms of word learning, Raboyeau et al. (2004) found that training French subjects to name English words (without focusing on any particular speech sounds or phonetic features) resulted in increased activation in the anterior cingulated cortex, left insular cortex, and right cerebellum. It is noteworthy that success in novel sound learning does not always imply the ability to use newly learned sounds in words. For example, Stager and Werker (1997) found that although 14-month-old infants could discriminate a minimal pair contrasting /b/ and /d/, they failed to learn to use these phonetically similar sounds in a word–object association task. In adult sound-to-word learning, Curtin et al. (1998) found that subjects' error pattern associated with perceiving a newly learned phonetic contrast depended on whether the experimental task required what they called "lexical access." In addition, Samuel (2002) argued that lexical knowledge can influence phoneme identification. Given this and the widely accepted notion of phonemes being building blocks of spoken languages, an understanding of the bridge between sound to word and their learning is essential for a fuller understanding of language processing and learning.

Much research in speech processing focuses on English and other Indo-European languages, especially the processing of consonants and vowels (segments). However, most languages of the world also use pitch patterns (called "lexical tones" or suprasegmentals) to mark word meaning (Fromkin, 2000). Mandarin Chinese is an example of a tone language that has four lexical tones: high-level (Tone 1), rising (Tone 2), dipping (Tone 3), and falling (Tone 4). Four different words can result when, for example, the syllable /ma/ is spoken in each of the four lexical tones. Respectively, it can mean "mother," "hemp," "horse," or "scold." Several recent crosslinguistic studies showed increased activation in the left inferior frontal regions (IFG) when pitch is used lexically, while the right homologous areas were activated during nonlexical pitch processing (e.g., Wong et al., 2004a; Gandour et al., 2003; Klein et al., 2001), suggesting that language experience and context affect brain responses. Wang et al. (2003) further found that training English speakers to identify Mandarin pitch patterns resulted in increased activation in the left superior temporal gyrus (STG) and right inferior frontal gyrus (IFG). It is worth noting that although the subjects in Wang et al. were first-year students of Mandarin Chinese, they were trained in the laboratory for two weeks to specifically identify pitch patterns for the purposes of the study. In other words, these subjects were not learning to use lexical tones in true lexical contexts. The right IFG results may be an indicator of the nonlinguistic nature of the training protocol, for Zatorre et al. (1992) and Wong et al. (2004a) found increased activation close to the right IFG when subjects performed nonlinguistic/nonlexical pitch processing. It remains unclear whether and what kind of neural changes occur when English-speaking subjects learn to use lexical tones, that is, not just categorize pitch patterns (Wang et al., 2003), but actually use them in words.

Sound-to-Word Learning

As discussed, second language learning studies, both behavioral and neural, largely focus either on the learning of foreign sounds without considering how they can be used in words, or on the learning of words without considering the contribution of the foreign phonetic features. In terms of behavioral studies, we are aware of one that examined the learning of a foreign segmental contrast (Thai voicing and stop consonants) and its contribution to word learning (Curtin et al., 1998). In terms of learning suprasegmental contrasts in words (phonetic features that are not associated with consonants or vowels and occur in most languages of the world), we have recently developed a training program investigating the learning of pitch patterns in word identification by native English-speaking adults (Wong and Perrachione, in press). Native English-speaking subjects who had no previous exposure to any tone languages learned six English pseudosyllables superimposed with three pitch patterns minimally (18 words) paired with pictures. For example, the subjects heard the syllable /pesh1/ ("pesh" spoken with a high level tone) while looking at a picture of a glass. When they saw the picture of a pencil, they heard the syllable /pesh2/ ("pesh" spoken with a rising tone). In other words, successful learning of the vocabulary necessarily entailed learning to use pitch in words. The goal of the vocabulary training was for subjects to learn to identify the picture of the appropriate object (e.g. a dog) when presented with a paired auditory stimulus (e.g. the syllable "ner4"); subjects were not required to produce the words they heard.

Subjects were trained with feedback three to four sessions per week and were tested on their knowledge of words without feedback at the end of each training session. Their performance on this last word identification test was used to determine whether the training criterion was met. In an attempt to observe ultimate attainment (learning), we trained subjects until their individual

TABLE I. Subjects were trained on a vocabulary of 18 artificial words (also used in the fMRI experiment)

pesh1 "glass"	dree1 "arm"	ner1 "boat"	vece1 "hat"	nuck1 "brush"	fute1 "shoe"
pesh2 "pencil"	dree2 "phone"	ner2 "potato"	vece2 "tape"	nuck2 "tissue"	fute2 "book"
pesh4 "table"	dree4 "cow"	ner4 "dog"	vece4 "piano"	nuck4 "bus"	fute4 "knife"

Word meanings assigned to the stimuli (listed in Table I) were high frequency English nouns (Raymer et al., Unpublished test). Each word is followed by its corresponding meaning in quotes. Numbers following the lexical items designate tone. Level tone is indicated by 1, rising tone by 2, and falling tone by 4, according to convention.

asymptotic performance was reached, defined by either an accuracy level in word identification of 95% or above for two consecutive sessions ("successful learning"), or when there was not a 5% improvement or better for four consecutive sessions ("less successful learning"). Attainment level is defined by accuracy in word identification in the first session at which asymptotic performance was shown. We found that while all subjects improved in their word identification, only about 53% of all subjects (9 out of 17) became successful learners. It took subjects on average about eight sessions to reach termination criterion with no significant difference between the successful and less successful learner groups.

In the present study, we investigated the neural changes associated with such learning. Subjects who participated in the pitch-to-word learning program, as described earlier, also participated in an identical fMRI experiment before and after training. Following previous crosslinguistic neuroimaging studies concerning lexical tones (e.g., Gandour et al., 2003; Klein et al., 2001; Wong et al., 2004a), we asked subjects to discriminate pitch patterns of the words they learned while brain blood oxygenation levels were measured. This task has been found to robustly distinguish individuals who spoke a tone language and those who did not (e.g., left lateralized activation for tone language subjects and right lateralized activation for nontone language subjects). Note that before training, such pitch patterns did not carry any lexical function for our current subjects, whereas after training, these pitch patterns became lexically meaningful. On the basis of other language learning studies (e.g., Golestani and Zatorre, 2004; McLaughlin et al., 2004), we expected significant individual differences in learning. We hypothesized that successful learning will show activation of a streamlined cortical network associated with language and linguistic pitch processing, including the left superior temporal (STG) and inferior frontal regions (e.g., Wong et al., 2003, 2004a), while asymptotically low performance will result in a diffused network of activation, including general attentional areas (e.g., prefrontal cortex) and nonlinguistic pitch processing areas (right STG and IFG).

MATERIALS AND METHODS

Subjects

Subjects were 17 young adult native speakers of American English (ages: 18–26 years, mean = 20.65) who reported having no audiologic and neurologic deficits. All were undergraduate students at, or recent graduates of, Northwestern University. All but one subject were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971); the remaining subject was ambidextrous. None of the subjects had previous exposure to a tone language at any time in life. They were the subjects in our behavioral training study discussed in the Introduction (Wong and Perrachione, in press). Appendix A includes basic demographic information as well as learning performance for each subject. As shown in Appendix A, there were two groups of subjects who did not differ in age or handedness scores. They were classified based on our definition of successful and less successful learners discussed in the Introduction.

fMRI Stimuli and Experimental Procedures

There were two sets of fMRI stimuli. The first set was identical to the training stimuli. These stimuli consisted of 18 English pseudowords with pitch (fundamental frequency) patterns resembling Mandarin Tones 1 (level), 2 (rising), and 4 (falling). As shown in Table I, there are six sets of words with minimal pitch contrasts in each set. The six base syllables (pesh, dree, ner, vece, nuck, and fute) were originally produced by a native speaker of American English and were subsequently resynthesized to include variants consisting of the three different pitch patterns using the Pitch Synchronous Overlap and Add method implemented in the software Praat (Boersman and Weeknik, 2005). These pitch contours implemented in the stimuli were modeled on the values obtained by Shih (1988), and the procedures of stimulus generation were similar to Wong et al. (2004a). Other than fundamental frequency (F0), all acoustic parameters corresponded to the talker's original productions, including duration and voice quality characteristics, so that each triad of the training stimuli differed only in F0. Eight native Mandarin-speaking individuals were asked to identify the pitch patterns of these training stimuli and performed at above 97% accuracy; these subjects also judged these stimuli to be perceptually natural.

Similar to our previous studies (Wong et al., 2003), the second set of stimuli consisted of time-varying sinusoids generated based on the first set; these served as stimuli for the control condition. Time-varying sinusoids were generated by using similar procedures: F0s of the word stimuli described were extracted at 10 ms intervals by using Wave-

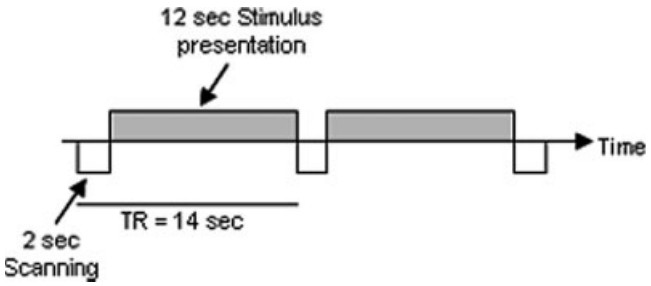

Figure 1.
Imaging sequence. Subjects were imaged at the first 2 s of a TR followed by 12 s (6 stimulus pairs) of stimulus presentations when no scanning took place.

surfer (Sjölander and Beskow, 2004), and time-varying sinusoids were then generated based on the extracted values using the synthesis software Tone (Tice and Carrell, 1997). Note that these stimuli do not carry any lexical function because they do not occur in words. The use of pitch stimuli with no lexical function as control stimuli is common in other lexical tone studies (e.g., Gandour et al., 1998).

There were two experimental conditions (fMRI runs) with the presentation order pseudorandomized. Within each run, subjects listened to pairs of stimuli in blocks of 6 pairs via headphones. For each run, there were 60 blocks of stimuli. Each pair of presentation, including response time, was fixed to 2 s. Each block, consisting of 6 pairs, lasted for 12 s. In the first run (Speech Condition), subjects listened to pairs of words from training and were asked to make a same-different judgment regarding the pitch (tone) patterns of the stimuli. Half of the stimulus pairs consisted of the same pitch patterns, and half consisted of different patterns. Within each pair, the segmental information (consonant-vowel sequence) was different, thus encouraging the subjects to rely on the pitch pattern for making judgments. Similar to our previous studies (Wong et al., 2003, 2004a), subjects were asked to press the appropriate button on a Cedrus response box to indicate their same-different judgments. The second (control) run (Sinewave Condition) was similar to the first run except that the stimuli consisted of time-varying sinusoids. These procedures were very similar to our previous studies investigating tone perception by native and non-native Mandarin speakers and learners (Wong et al., 2003, 2004a).

MRI Acquisition and Analyses

Functional and anatomical MR images were acquired at the Center for Advanced Magnetic Resonance Imaging (CAMRI) at the Northwestern University Department of Radiology using a Siemens 3 Tesla Trio whole body machine. For each subject, a high resolution, T1-weighted 3D volume was acquired (MP-RAGE with a TR/TE of 2100 ms/2.4 ms, flip angle of 15°, TI of 1100 ms, matrix size of 256 × 256, FOV of 22 cm, slice thickness of 1 mm). The orientation of this 3D volume was sagittal. The 3D volume was used in conjunction with the activation maps to localize the function and to determine the anatomic regions for investigation of the time course data. The T2*-weighted functional images were acquired axially by using a susceptibility weighted EPI pulse sequence while the subjects performed behavioral tasks. A TE of 30 ms, a TR of 14 s, a flip angle of 90°, and a slice thickness of 3 mm were prescribed. A 14-s TR (sparse sampling method) allowed for the scanner to be quiet during stimulus presentation and thus minimized contamination of the acoustic stimuli (e.g., Belin et al., 1999; Hall et al., 1999; Wong et al., 2005). Images were taken at the first two seconds of each TR (no stimuli was presented during this 2-s period). Each block of stimulus presentation was 12 s long and was between two consecutive functional scans. Figure 1 shows an example of the sequence of stimulus presentation and image acquisition. For each run, subjects were presented 60 blocks of stimuli and 20 blocks of silence randomly inserted (80 blocks total; 1120 s or 18 min 40 s. The T2*-weighted functional MR images (time series) were sampled at 3 × 3 × 3 mm voxel dimensions and were analyzed by BrainVoyager (Goebel, 2004). The data were first linearly detrended, motion and scan time corrected, and spatially smoothed (FWHM 6 mm). After these preprocessing procedures, hemodynamic responses were estimated. Square waves modeling the events of interest were created as extrinsic model waveforms of the task-related hemodynamic response. These events of interest included the two experimental conditions before and after training (4 events total). Note that even though the TR was 14 s long, image acquisition only occurred during the first two seconds of the TR as stated, as opposed to the entire TR. Thus, the images collected reflected either a stimulus event occurring at one of these time

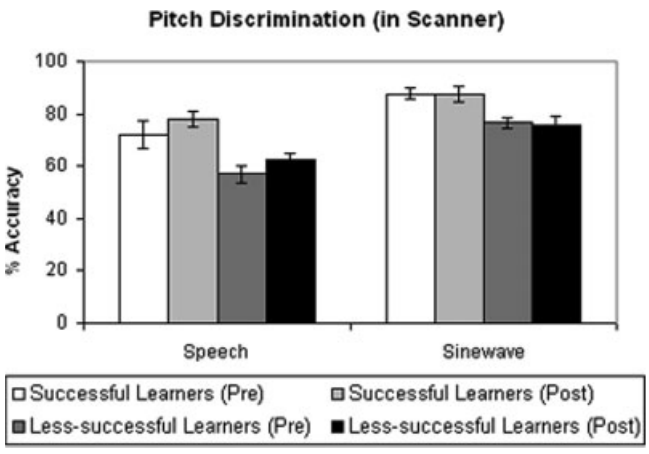

Figure 2.
Subjects' pitch discrimination performance (in the scanner). Error bars show standard errors of the mean.

TABLE II. Regions of activation

Area (BA)	*X*	*Y*	*Z*	*t* value	Size (mm^3)
Pre-training speech vs. Pre-training sinewave (All subjects)					
R STG (BA 41/42)	57	−25	10	4.63	494
L STG (BA 22)	−63	−16	4	4.91	267
Post-training speech vs. Post-training sinewave (All subjects)					
R STG (BA 22)	42	−22	4	5.44	3480
L STG (BA 22)	−45	−13	4	5.41	10226
L PFC (∼BA 9/46)	−39	17	25	4.81	7707
L ITG (BA 37)	−53	−43	−14	4.65	664
L IFG (BA 45)	−60	20	10	3.95	466
L IPL (BA 40)	−54	−34	31	4.02	129
Post-training speech vs. Pre-training speech (All subjects)					
R STG (BA 22)	66	−13	−2	4.94	293
R STG (BA 21)	63	−37	−8	5.52	3934
R ITG (BA 20)	39	−16	−20	5.18	3024
R pITG (BA 19/37)	48	−64	−15	4.05	328
R PFC (BA 6)	46	−4	34	4.69	1339
R PFC (∼BA 46)	36	23	25	4.39	241
R Cuneus (BA 18)	12	−73	25	4.15	200
R Caudate Body	11	5	8	3.69	130
L ParahipG	−24	−43	−2	5.69	1166
L Putamen	−21	−4	7	4.53	636
L PFC/IFG (BA 6)	−42	−4	25	4.81	3093
L ITG (BA 20/36/37)	−30	−34	−14	3.92	179
L STG/MTG (BA 41/42)	−54	−19	10	5.40	5576
L Precuneus (BA 19)	−33	−67	37	4.24	208
L Postcent (BA 2)	−39	−37	61	4.74	676
L pMTG (BA 22)	−48	−43	1	4.02	133
L IFG (BA 44)	−51	11	10	4.54	386
L pITG (BA 37/20)	−58	−50	−10	6.45	2703
Succeesful vs. Less successful learners (Post-training speech condition)					
Successful > Less successful learners					
L pSTG (BA 22)	−60	−40	7	5.16	436
L TTG (BA 41)	−39	−34	10	5.31	157
Less successful > Successful learners					
R MTG/STS (BA 21)	54	−28	−8	−4.14	145
R ITG (∼BA 20)	45	−22	−20	−4.98	756
R PFC/Precent (BA 6)	48	−4	25	−4.01	287
R IFG (∼BA 46/10)	30	38	13	−4.98	2247
R mSTG/Insula (BA 21)	42	−7	−8	−4.21	460
R IFG (∼BA 47)	39	42	−10	−4.92	225
R aSTG (BA 38/47)	39	11	−20	−4.63	385
R meSTG/ParahipG (BA 19/37)	33	−43	−5	−4.13	487
R Putamen	21	11	22	−3.82	182
R MeFG (BA 10/32)	21	47	10	−4.30	1457
R Ant Cing (BA 32)	9	26	25	−4.07	624
R Caudate Head	9	2	4	−4.12	1546
R MeFG (BA 9)	9	47	34	−4.50	407
L MeFG (BA 10)	−6	53	4	−4.80	1318
R Precuneus/SPL (BA 7)	3	−67	34	−3.71	137
L MeFG (∼BA 7)	−21	38	25	−3.93	251
L PFC/MFG (BA 9)	−36	26	31	−5.96	3601
L Precent (∼BA 6)	−32	−4	40	−4.24	200
L ParahipG (BA 19)	−33	−43	−2	−4.38	392
L PFC/IFG (∼BA 9)	−39	5	23	−4.35	1298
L Ant Insula (∼BA 13/ ∼BA 45)	−33	14	4	−4.41	348
L Post Insula/STG (BA 13)	−39	−16	7	−4.84	688
L aMTGi (BA 21)	−48	−1	−20	−5.02	1484
L pITG (BA 20)	−51	−31	−14	−4.77	437
L IPL/Supramarg (BA 40)	−54	−46	34	−5.06	324
L Postcent (BA 2)	−51	−19	31	−4.43	193
L Postcenti (BA 43)	−52	−16	13	−4.27	192

TABLE II. (continued)

Area (BA)	X	Y	Z	t value	Size (mm^3)
Successful vs. Less successful learners (Pre-training speech condition)					
Successful > Less successful learners					
R STG/MTG (BA 21/22)	48	−28	−2	6.54	5790
R pMTG (BA 39)	54	−58	7	3.96	301
R pITG/Fusi (BA 37)	42	−61	−11	5.69	3413
R IOG/LingG (BA 17)	15	−94	−12	4.56	670
Corpus Coll	3	−22	25	4.23	161
L MOG (BA 18)	−24	−82	−8	5.91	3509
L STG (BA 22)	−60	−37	7	7.46	3764
L pMTG/OG (BA 37/19)	−51	−61	−8	4.85	209
L pSTG (BA 37)	−54	−64	4	5.66	656
Less Successful > Successful learners					
R STG (BA 22/∼42)	42	38	4	−4.10	151
R MFG/∼IFG (BA 47)	36	49	22	−5.94	996
R ParahipG	30	−10	−20	−4.35	148
R MeFG (BA 10)	18	44	4	−4.64	639
R MeFG (BA 10)	3	56	1	−5.09	617
R Ant Cing (BA 24)	2	32	−5	−4.74	428
Successful vs. Less successful learners (Post- vs. Pre-training speech condition)					
Less Successful > Successful learners					
R STS (∼BA 21)	54	−25	−5	−7.17	1171
R ParahipG (BA19)[a]	30	−43	−5	−5.24	3400
R Insula (BA 13)	27	−31	19	−3.96	126
R Precuneus (BA 7)	2	−67	31	−4.00	141
L Insula (BA13)	−31	−10	22	−3.87	148
L MeFG	−18	−28	31	−4.33	174
L MOG (BA 18)	−24	−85	−8	−4.85	1000
L PFC/MFG (BA 9)	−39	26	31	−4.30	142
L mSTG (∼BA 42/41)	−57	−22	7	−5.04	999
L pITG (BA 37)	−54	−58	−8	−4.71	198

The coordinates represent the location of the peak voxel for a cluster in Talaraich space. BA: Brodmann's Area within a 3 mm radius, or ∼ indicates the closest BA outside 3 mm. L: Left, R: Right, STG: superior temporal gyrus, aSTG: anterior superior temporal gyrus, mSTG: middle superior temporal gyrus, pSTG: posterior superior temporal gyrus, meSTG: medial superior temporal gyrus, STS: superior temporal sulcus, aMTG: anterior middle temporal gyrus, pMTG: posterior middle temporal gyrus, TTG: transverse temporal gyrus (Heschl's Gyrus), ITG: inferior temporal gyrus, pITG: posterior inferior temporal gyrus, Fusi: fusiform gyrus, IFG: inferior frontal gyrus, PFC: prefrontal cortex, MFG: middle frontal gyrus, MeFG: medial frontal gyrus, Postcent: postcentral gyrus, Postcenti: inferior postcentral gyrus, Ant Insula: anterior insula, Post Insula: posterior insula, IPL: inferior parietal lobule, Supramarg: supramarginal gyrus, SPL: superior parietal lobule, MOG: middle occipital gyrus, ParahipG: parahippocampal gyrus, Ant Cing: anterior cingulate, LingG: lingual gyrus, Corpus Coll: corpus callosum.

[a] Large cluster extending to MTG.

points or a null event (no stimulus presentation). Imaging at specific time points removed the need to convolve the task-related extrinsic waveforms with a hemodynamic response function before statistical analyses, as is commonly done. The waveforms of the modeled events were used as regressors in a multiple linear regression of the voxel-based time series. Beta values signifying the fit of the regressors to the functional scanning series, voxel-by-voxel, for each condition, were obtained for each subject. Anatomical and functional images from each subject were transformed into the Talairach stereotaxic space (Talairach and Tournoux, 1988) at $1 \times 1 \times 1$ mm voxel dimensions for both sets of images. Beta values were normalized before entering into a multisubject general linear model, including contrasting the hemodynamic responses associated with the different runs before and after training.

RESULTS

Behavioral Results (From fMRI Sessions)

All subjects discriminated pitch patterns in words (Speech Condition) and time-varying sinusoids without the syllables (Sinewave Condition) significantly above chance level. These behavioral performances are comparable with our previous study using similar discrimination tasks by native English- and Mandarin-speaking subjects (Wong et al., 2004a). Accuracy data were entered into a $2 \times 2 \times 2$ repeated measures ANOVA (group $\times$ condition $\times$ training). We found a main effect of condition [F (1, 1, 15) = 53.87, $p < 0.0001$], showing discrimination accuracy in the Sinewave Condition to be higher than the Speech Condition for all subjects. We also found a significant training $\times$ task interaction [F (1, 1, 15) = 10.38, $p < 0.006$],

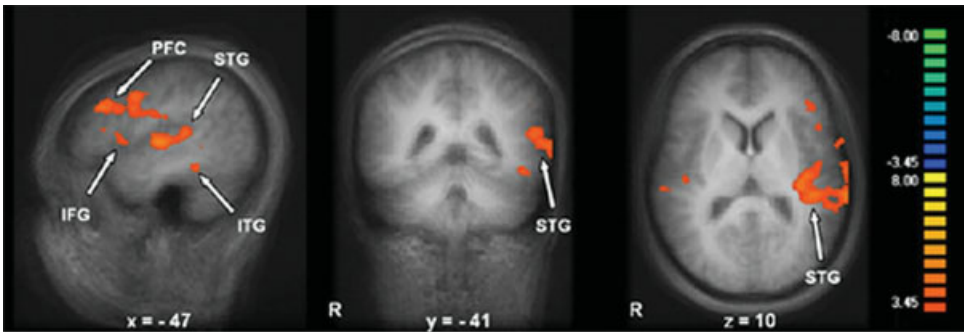

Figure 3.
Brain activation revealed by the *Post-Training Speech vs. Sinewave* (*All Subjects*) contrast. For this and subsequent figures, activation is superimposed on an averaged T1-weighted volume (in Talairach space). Color scale represents normalized *t* value and applies to this and subsequent figures.

showing that subjects' performance in the Speech Condition improved more so than the Sinewave Condition after training. The main effect of group (successful learners better than less successful learners) was only marginally reliable [F (1, 1, 15) = 2.07, $p < 0.17$]. These results are summarized in Figure 2. There are no other main effects or significant interactions.

Imaging Results

Imaging results are classified into two parts: voxel-wise contrasts and region-of-interests (ROI) analyses.

Voxel-wise contrasts

We report several voxel-wise contrasts in this section. Only clusters exceeding an uncorrected single-voxel p value of <0.0005 ($t = 3.45$) extending at least 125 mm^3 were reported. Table II shows results of the various contrasts including the t value of the peak voxel of a given cluster as well as its size.

Pre-training speech vs. pre-training sinewave (all subjects). Before training, subjects (from both groups) showed increased activation in the superior temporal region bilaterally when listening to speech relative to sinewave stimuli. These results replicated Wong et al. (2003) when Mandarin learners of low proficiency performed the same tasks.

Post-training speech vs. post-training sinewave (all subjects). After training, subjects also showed increased activation in the superior temporal region bilaterally when listening to speech relative to sinewave stimuli; however, activation in the STG was more extensive spatially. In addition, activation in the left prefrontal cortex (PFC), inferior temporal lobe (ITG), IFG, and inferior parietal lobule (IPL) was also noted. Figure 3 highlights some of the results.

Post-training speech vs. pre-training speech (all subjects). When directly comparing activation in the Speech Condition after training relative to before training, we found increased activation in bilateral STG, ITG, parietal, and basal ganglia regions. In addition, activation in the left IFG (BA 44) was observed.

Successful vs. less successful learners (post-training speech condition). The aforementioned contrasts considered activation before and after training when the two groups of subjects were combined. The following contrasts focus on comparing the two subject groups. When comparing successful and less successful subjects in the post-training Speech Condition, we found that successful learners showed increased activation in the left posterior superior temporal gyrus (pSTG) (see Fig. 4) in addition to a smaller activating cluster in the left transverse temporal gyrus (TTG). The posterior STG activation is 8 mm posterior to the most posterior point of TTG (BA 41; y = −32) on the same axial plane (z axis). On the other hand, relative to the successful learners, the less successful learners showed increased activation in numerous areas, including the right

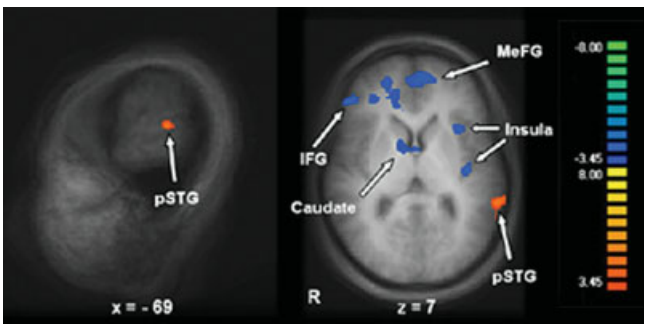

Figure 4.
Brain activation revealed by the *Successful vs. Less Successful Learners* (*Post-Training Speech*) contrast. Stronger activation for the successful and less successful learners are indicated by red and blue clusters, respectively.

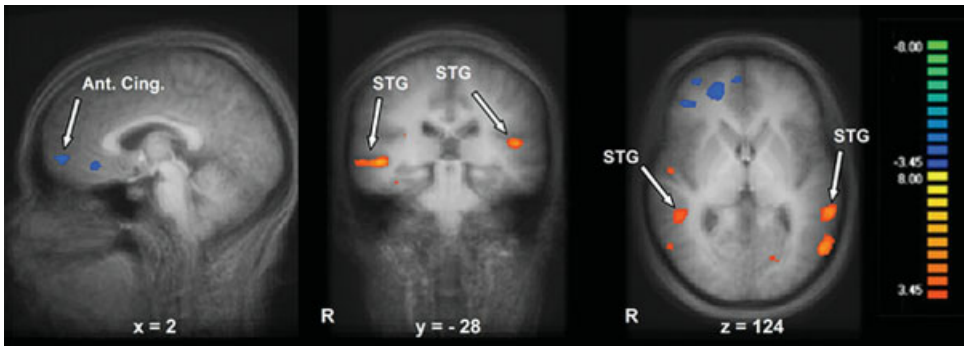

Figure 5.
Brain activation revealed by the *Successful vs. Less Successful Learners* (*Pre-Training Speech*) contrast.

superior temporal gyrus and sulcus (but not left), and bilateral medial frontal, prefrontal, inferior temporal, and parietal regions. In addition, there was increased activation in the right IFG (BA 46 and 47).

Successful vs. less successful learners (pre-training speech condition). Although differences in brain activation were observed after training between the two subject groups, there were also differences between the two groups even before training. Relative to the less successful learners, successful learners showed increased activation in the bilateral superior and middle temporal regions, as well as the right ITG. The less successful learners showed increased activation in the right medial frontal, anterior cingulate, and middle frontal (close to BA 47) areas, as well as a small cluster of activation in the right superior temporal region close to planum temporale (BA 42). Some of these results are highlighted in Figure 5.

Successful vs. less successful learners (post-training vs. pre-training speech condition). In addition to considering differences between the two subject groups at one time point (before or after training), we also considered the differences in changes (i.e., post- vs. pre-training) between the two groups. We found no areas that showed more changes in the successful relative to the less successful group. On the contrary, relative to the successful learners, the less successful learners showed additional changes in many brain regions, the strongest of which being in the right superior temporal sulcus. Changes were observed in the left STG, but in an area much more anterior to the posterior region found in the *Successful vs. Less Successful* (*Post-training Speech*) contrast discussed earlier (see Fig. 4). Increased activation in the left PFC and ITG was also found.

Region-of-interest analyses

The voxel-wise contrasts reported above suggest that not only did the successful and less successful subjects differ in their brain activation post-training, differences were observed before training. To further investigate these differences, we performed the following region-of-interest (ROI) analysis. We identified the strongest 5 × 5 × 5 activating cluster in the superior temporal gyri in both hemispheres based on the *Post-Training Speech vs. Pre-Training Speech* contrast from all subjects as described earlier. This cluster can be viewed as the cluster showing the strongest change between post- and pre-training sessions across subjects. For each subject, we then calculated the percent signal change for all experimental conditions (Speech and Sinewave, before and after training) in this region and entered them into a 2 × 2 × 2 (subject group × task × training) repeated measures ANOVA. Figure 6 shows the results from the left superior temporal gyrus. For clarity, we separated these results into a pre-training (top panel)

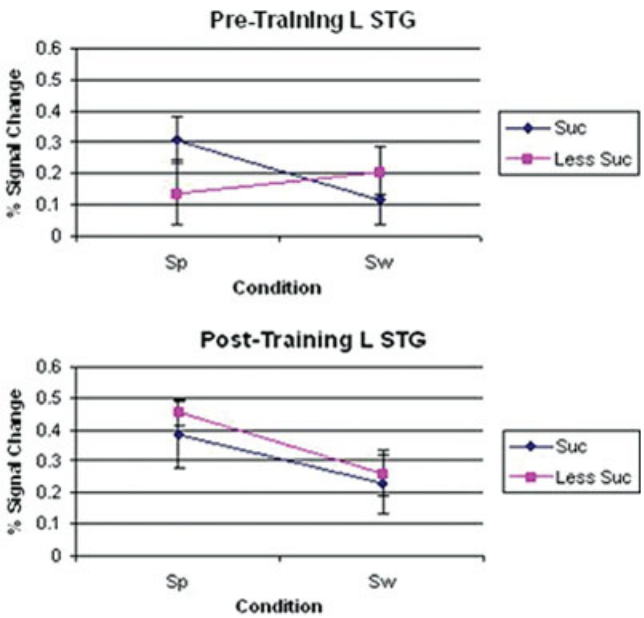

Figure 6.
Left STG activation (in the strongest activating 5 mm^3 cluster) in Speech (Sp) and Sinewave (Sw) conditions before (top) and after (bottom) training. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

and post-training figure (bottom panel). For the left hemisphere, we found a marginal main effect of training [F (1, 15) = 3.806, $p < 0.07$] and a main effect of task [F (1, 15) = 6.650, $p < 0.02$], indicating increased activation after training and increased activation in the Speech Condition. We found a significant training × task interaction [F (1, 15) = 5.299, $p < 0.03$], indicating a larger increase in activation in the Speech Condition post-training. Most interestingly, we also found a significant training × task × group [F (1, 15) = 6.114, $p < 0.02$] interaction, indicating a decreased activation in the Speech Condition before training in the less successful subject group relative to their successful counterparts, confirming results from the *Successful vs. Less Successful Learners (Pre-Training Speech Condition)* contrast. The results for the right hemisphere were similar, but with the absence of training × task interaction, indicating similar increases in activation in both Speech and Sinewave conditions after training. Specifically, we found a main effect of training [F (1, 15) = 4.476, $p < 0.052$], main effect of task [F (1, 15) = 5.112, $p < 0.04$], and a marginally significant training × task × group interaction [F (1, 15) = 4.166, $p < 0.059$].

DISCUSSION

Although there has been a wealth of recent studies investigating the neural correlates of pitch (e.g., Schneider et al., 2005) and speech perception (e.g., Binder et al., 2004; Wong et al., 2004b) in adults, our study is among the first to investigate learning involving speech perception and to consider the contribution of foreign speech sounds in word learning. This study is complementary to studies investigating the learning of foreign speech sounds in nonlexical contexts (e.g. Callan et al., 2003; Golestani and Zatorre, 2004) and the learning of foreign words when phonetic details are irrelevant (McLaughlin et al., 2004). This study is also complementary to Wang et al. (2003) concerning English speakers learning to identify Mandarin pitch patterns. Although Wang et al. found that training English speakers to identify Mandarin pitch patterns resulted in increased activation in the left STG and right IFG, the subjects were trained in the laboratory for two weeks to specifically identify pitch patterns for the purpose of the study. In other words, these subjects were not learning to use lexical tones in true lexical contexts. The right IFG results may be an indicator of the nonlinguistic nature of the training protocol (Wong et al., 2004a; Zatorre et al., 1992).

In the present study, we found that learning to use pitch patterns in words by English-speaking individuals resulted in changes in a network of brain regions involving association auditory cortex, inferior frontal gyrus, prefrontal cortex, basal ganglia, medial frontal cortex, medial temporal cortex, posterior parietal cortex, and inferior temporal lobe. Most interestingly, depending on the subjects' attainment level, not all of these areas were activated. Specifically, successful learning is characterized by notable streamlined activation in the left superior temporal region, most notably pSTG, after training. Less successful learning is characterized by a diffused network of activation that includes the right superior temporal and right inferior frontal regions, which have been implicated in previous studies using a similar behavioral task involving nonlinguistic pitch perception (e.g., Gandour et al., 1998; Wong et al., 2003, 2004), prefrontal cortex, medial frontal cortex, and posterior parietal regions, which have been shown to be associated with increased working memory, attention, and effort (see Kelly & Garavan (2004) for a review), and basal ganglia and parahippocampal gyrus, which have been implicated in learning involving procedural and declarative memory (Ullman, 2004). Although the left STG was also found to be activated in the less successful learner group, its location is in mid-STG rather than pSTG as found in the successful learning group.

It has been proposed that the auditory cortex can be separated into two streams—a dorsal "where" stream specialized in auditory spatial processing and a ventral "what" stream specialized in auditory object recognition (Rauschecker, 1998). More specific to speech, Hickok and Poeppel (2004) as well as Scott and Wise (2004) proposed that the ventral stream is especially important for the processing of intelligible speech sounds (e.g., Liebenthal et al., 2005; Scott et al., 2000), while the dorsal stream is important for motor-speech integration (e.g., Buchsbaum et al., 2001; Wise et al., 2001). Although Scott and Wise (2004) explicitly predict the dorsal stream to be important for language acquisition, especially in relation to the possible involvement of covert and overt rehearsal of words during learning, little empirical evidence exists implicating the role of the dorsal stream in speech-to-word learning. Our results are among the first to demonstrate the role of the posterior (dorsal) auditory neural stream in successful speech-to-word learning. This pattern of activation is consistent with anecdotal reports from our learners who credited vocal rehearsal to be important to their success in learning. In the proposal by Hickok and Poeppel, after speech sounds are encoded, word meaning is further processed in the posterior inferior temporal lobe (pITG). As shown in the post-training vs. pre-training Speech Condition contrast, subjects, regardless of group membership, showed increased activation in the inferior temporal region after training. This is especially true for the less successful learners as revealed by the contrast comparing the two learner groups in the post-training Speech Condition. The functional task (Speech Condition) in the current study involved the discrimination of an aspect of phonology that did not require direct access of meaning. However, because subjects were trained on associating the stimuli with meaning, it is possible that they were unable to completely ignore meaning. The less successful subjects' reduced performance in the Speech Condition may indicate interference by the word meaning, which in turn resulted in greater activation in the ITG in this group.

Besides the kind of learning associated with speech and language, there exists a body of literature on other types of learning and practice, which provides converging evidence on how the nervous system handles novel information and acquires new skills, such as piano playing and practice (Bengtsson et al., 2005; Hasegawa et al., 2004) and music reading (Stewart et al., 2003). Of note, Little and Thulborn (2005) trained subjects to classify random dots into categories and found that as behavioral performance improved, a reduction of brain activation in visuospatial and spatial attention brain regions was observed, suggesting an increase in neural efficiency. Furthermore, Booth et al. (2001) found activation of unimodal and heteromodal cortical areas in processing word forms in adults and children respectively, suggesting streamlining during development. These results are largely consistent with our findings that successful learners show distinct areas of activation while the less successful learners show increased activation in a diffused brain network.

Kelly and Garavan (2004) reviewed practice-related neural changes studies and concluded that there are at least two types of activation patterns: true reorganization and redistribution (pseudo-reorganization). A true reorganization is associated with subjects' performance being cognitively and neurobiologically different before and after training; namely, the location of brain activation is changed. A redistribution is associated with practice resulting in attainment of automatic performance, rather than a true cognitive shift. Areas of activation show increase or decrease activations, though the same areas are involved before and after training. Most importantly, a redistribution often involves "generic attentional and control areas" (Kelly and Garavan, 2004) or a "scaffolding" network with novel task demand (Petersen et al., 1998), including prefrontal cortex, anterior cingulate cortex, and posterior parietal cortex, especially in early stages of learning. The less successful learners in the present study clearly showed neural characteristics consistent with the early stages of redistribution, as revealed by the activation of a "generic" attentional network. However, it is unclear whether the successful learners underwent true reorganization. It is true that they showed increased activation in the left STG after training; however, their left STG was strongly activated in the successful learners group even before training relative to the less successful learners. As discussed in greater detail in Wong and Perrachione (in press), these successful learners were also more likely to have received extensive formal musical training (at least 5 years in one instrument started before the age of 10 years old). This might have contributed to how pitch is organized in their brains even before training, as musicians and nonmusicians showed different brain mechanisms in pitch processing (Gaab and Schlaug, 2003).

A marked cerebral lateralization distinction is often found in studies of lexical tone perception involving subjects who were native (left lateralization) and non-native speakers (right lateralization) of a tone language (e.g., Gandour et al., 1998; Wong et al., 2004a). In the present study, we found bilateral, rather than right lateralized, STG activation in our subjects before training. It is noteworthy that unlike the present study, none of these previous studies used time-varying sinusoids in the control condition. Before training, a main difference between the sets of stimuli in the experimental ("Speech") and control ("Sinewave") conditions was that the former was broader in bandwidth. It has been found that STG (lateral auditory cortex) is sensitive to broadband acoustic signals (Rauschecker, 1998). Since neither sets of stimuli carried lexical function before training, the brain differences observed were likely related to differences in acoustic complexity rather than differences in functional status (lexical vs. nonlexical). Because the stimuli were presented binaurally, it is not surprising that the lateral auditory cortex in both hemispheres was activated. After training, however, the experimental stimuli carried a lexical function; thus, a left lateralization pattern, driven by functional differences, was observed.

Research studies concerning learning in adulthood often show or assume that little behavioral and neurological differences exist before training among different subject groups. Contrary to this assumption, the present study showed pre-training differences between subject groups. Such information before training, together with a host of other factors, can potentially be used in the future for the placement of learners into training programs that are most likely to be beneficial for language learning.

ACKNOWLEDGMENTS

The authors thank Jay Mittal, Carson Lam, Ann Bradlow, Gnyan Patel, Andrew Mazotas, Catherine Warrier, and Nondas Leloudas for their assistance in this research.

REFERENCES

Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B (1999): Event-related fMRI of the auditory cortex. Neuroimage 10:417–429.

Bengtsson SL, Nagy Z, Skare S, Forsman L, Forssberg H, Ullen F (2005): Extensive piano practicing has regionally specific effects on white matter development. Nat Neurosci 8:1148–1150.

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004): Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295–301.

Bley-Vroman R (1989): What is the logical problem of foreign language learning? In: Gass SM, Schachter J, editors. Linguistic Perspectives on Second Language Acquisition. Cambridge, UK: Cambridge University Press. pp 41–68.

Boersman P, Weenink D (2004): Praat, version 4.126.

Booth JR, Burman DD, Van Santen FW, Harasaki Y, Gitelman DR, Parrish TB, Mesulam MM (2001): The development of specialized brain systems for reading and oral-language. Child Neuropsychol 7:119–141.

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

Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R (2003): Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. Neuroimage 19: 113–124.

Curtin S, Goad H, Pater JV (1998): Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. Second Lang Res 14:389–405.

Fromkin VA (2000): Linguistics: An Introduction to Linguistic Theory. Oxford: Blackwell.

Gaab N, Schlaug, G (2003): The effect of musicianship on pitch memory in performance matched groups. Neuroreport 14: 2291–2295.

Gandour J, Wong D, Hutchins G (1998): Pitch processing in the human brain is influenced by language experience. Neuroreport 9:2115–2119.

Gandour J, Dzemidzic M, Wong D, Lowe M, Tong Y, Hsieh L, Satthamnuwong N, Lurito J (2003): Temporal integration of speech prosody is shaped by language experience: An fMRI study. Brain Lang 84:318–336.

Goebel R (2004): BrainVoyager QX. Psychology Software Tools: Pittsburgh, PA.

Golestani N, Zatorre RJ (2004): Learning new sounds of speech: Reallocation of neural substrates. Neuroimage 21:494–506.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999): Sparse temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223.

Hasegawa T, Matsuki K, Ueno T, Maeda Y, Matsue Y, Konishi Y, Sadato N (2004): Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale: An fMRI study. Brain Res Cogn Brain Res 20: 510–518.

Hickok G, Poeppel D (2004): Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition 92:67–99.

Kelly AM, Garavan H (2004): Human functional neuroimaging of brain changes associated with practice. Cereb Cortex 15:1089–1102.

Klein D, Zatorre RJ, Milner B, Zhao V (2001): A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. Neuroimage 13:636–653.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005): Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631.

Little DM, Thulborn KR (2005): Correlations of cortical activation and behavior during the application of newly learned categories. Brain Res Cogn Brain Res 25:33–47.

McLaughlin J, Osterhout L, Kim A (2004): Neural correlates of second-language word learning: Minimal instruction produces rapid change. Nat Neurosci 7:703–704.

Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

Petersen SE, van Mier H, Fiez JA, Raichle ME (1998): The effects of practice on the functional anatomy of task performance. Proc Natl Acad Sci USA 95:853–860.

Raboyeau G, Marie N, Balduyck S, Gros H, Demonet JF, Cardebat D (2004): Lexical learning of the English language: A PET study in healthy French subjects. Neuroimage 22:1808–1818.

Rauschecker JP (1998): Cortical processing of complex sounds. Curr Opin Neurobiol 8:516–521.

Samuel AG (2001): Knowing a word affects the fundamental perception of the sounds within it. American Psychological Society 12:348–351.

Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C, Rupp A (2005): Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nat Neurosci 8:1241–1247.

Scott SK, Wise RJ (2004): The functional neuroanaomty of prelexical processing in speech perception. Cognition 92:13–45.

Scott SK, Blank CC, Rosen S, Wise RJ (2000): Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Selinker L (1972): Interlanguage. Int Rev Appl Linguist 10:209–231.

Shih C-L (1988) Tone and intonation in Mandarin. Work Pap Cornell Phon Lab 3:83–109.

Sjölander K, Beskow J (2004): Wavesurfer. Stockholm: Kungliga Tekniska Högskolan (Royal Institute of Technology).

Stager CL, Werker JF (1997): Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature 388:381, 382.

Stewart L, Henson R, Kampe K, Walsh V, Turner R, Frith U (2003): Brain changes after learning to read and play music. Neuroimage 20:71–83.

Talairach J, Tournoux P (1988): Co-planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging. New York: Thieme.

Tice B, Carrell T (1997): Tone. Speech Perception Laboratory, University of Nebraska: Lincoln, NE.

Ullman MT (2004): Contributions of memory circuits to language: The declarative/procedural model. Cognition 92:231–270.

Wang Y, Sereno JA, Jongman A, Hirsch J (2003): fMRI evidence for cortical modification during learning of Mandarin lexical tone. J Cogn Neurosci 15:1019–1027.

Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001): Separate neural sub-systems within 'Wernicke's' area. Brain 124:83–95.

Wong PCM, Perrachione TK (in press): Learning pitch patterns in lexical identification by native English-speaking adults. Applied Psycholinguistics.

Wong PCM, Nusbaum HC, Skipper J, Small SL (2003): Cortical activation associated with lexical tone acquisition. Presented at 10th Annual Meeting of the Cognitive Neuroscience Society, New York, March 30–April 1.

Wong PCM, Parson LM, Martinez M, Diehl RL (2004a): The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. J Neurosci 24:9153–9160.

Wong PCM, Nusbaum HC, Small SL (2004b): Neural bases of talker normalization. J Cogn Neurosci 16:1173–1184.

Wong PCM, Lee KM, Parrish TB (2005): Neural bases of listening to speech in noise. In Proceedings of Interspeech 2005—Eurospeech—9th European Conference on Speech Communication and Technology, Lisbon, September 4–8.

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

APPENDIX

Age in years; M = male, F = female, Handedness on a scale from -100 (left-handed) to 100 (right-handed); Initial and Attainment values are percent correct in the word identification behavioral (training) task

Subject #	Age	Sex	Handedness	Initial	Attainment
Successful learners					
E-05-010	21	M	57.14	53.70	98.15
E-05-016	20	M	86.67	55.56	100.00
E-05-052	19	F	100.00	37.04	94.44
E-05-057	20	M	100.00	27.78	96.30
E-05-060	20	F	0.00	51.85	96.30
E-05-062	25	F	75.00	16.67	96.30
E-05-064	19	F	100.00	20.37	96.30
E-05-082	23	F	100.00	31.48	98.15
Mu-05-016	21	F	80.00	35.19	98.15
Mean	20.89		77.65	36.63	97.12
Less successful learners					
E-05-055	21	F	88.89	20.37	85.19
E-05-061	21	F	85.71	25.93	53.70
E-05-063	18	M	86.67	11.11	90.74
E-05-068	20	M	71.43	31.48	57.41
E-05-070	18	F	84.62	22.22	57.41
E-05-076	26	M	100.00	33.33	57.40
E-05-086	18	F	100.00	38.89	42.60
E-05-097	21	M	40.00	35.19	96.3[a]
Mean	20.38		82.17	27.31	63.49

[a] Although this subject achieved a high accuracy during one session, this level of performance was not maintained, and as such the subject did not fall under our definition of "successful learner," set forth in the Introduction.