Volume of Left Heschl's Gyrus and Linguistic Pitch Learning

Patrick C.M. Wong[1,2], Catherine M. Warrier[1], Virginia B. Penhune[3,6], Anil K. Roy[1], Abdulmalek Sadehh[4], Todd B. Parrish[5] and Robert J. Zatorre[6,7]

[1]The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208, USA, [2]Northwestern University Institute for Neuroscience, Northwestern University, Evanston, Illinois, 60208, USA, [3]Department of Psychology, Concordia University, Montreal, Qubec, Canada, H4B1R6 [4]Department of Behavioral Medicine and Psychiatry, West Virginia University School of Medicine–Charleston Area Medical Center, Morgantown, West Virginia, 26505, USA, [5]Departments of Radiology and Biomedical Engineering, Northwestern University, Chicago, Illinois, 60611, USA, [6]Montreal Neurological Institute, Montreal, Quebec, Canada, H3A2B4, [7]Department of Neuropsychology, McGill University, Montréal, Quebec, Canada, H3A2B4

Research on the contributions of the human nervous system to language processing and learning has generally been focused on the association regions of the brain without considering the possible contribution of primary and adjacent sensory areas. We report a study examining the relationship between the anatomy of Heschl's Gyrus (HG), which includes predominately primary auditory areas and is often found to be associated with nonlinguistic pitch processing and language learning. Unlike English, most languages of the world use pitch patterns to signal word meaning. In the present study, native English-speaking adult subjects learned to incorporate foreign pitch patterns in word identification. Subjects who were less successful in learning showed a smaller HG volume on the left (especially gray matter volume), but not on the right, relative to learners who were successful. These results suggest that HG, typically shown to be associated with the processing of acoustic cues in nonspeech processing, is also involved in speech learning. These results also suggest that primary auditory regions may be important for encoding basic acoustic cues during the course of spoken language learning.

Keywords: auditory cortex, auditory perception, Heschl's Gyrus, language processing, speech learning, speech perception

Introduction

The human auditory system has the remarkable ability to incorporate complex acoustic signals into spoken language. Normal variations of this ability to learn in adulthood is often indicated by a wide range of successful patterns such that only a small number of individuals show native-like attainment level (e.g., Bongaerts 1999). Undoubtedly, successful learning is likely to be multifaceted and various behavioral factors have been identified, including verbal working memory (e.g., Miyake and Friedman 1998), motivation, age of onset, and length, intensity, and quality of training (e.g., Birdsong 1999, Bongaerts 1999). Less known is how brain anatomy contributes to the success of spoken language learning. An understanding of how preexisting neuroanatomic differences can have an impact on adult learning is not only theoretically interesting, as it informs us about brain organization and limits of plasticity, but it also has significant clinical implications as it can assist the development of optimal training/rehabilitation programs.

Extending our previous studies examining behavioral (Wong and Perrachione, forthcoming) and neurophysiologic (cerebral hemodynamic responses measured by functional magnetic resonance imaging [fMRI]; Wong et al., forthcoming) factors in the same group of subjects, the current study examines the association between brain structures and adult spoken word learning ability. Unlike English, most languages of the world, called tone languages, use pitch patterns (primarily signaled by fundamental frequency [F0]) to mark individual word meaning (Fromkin 2000). In Wong and Perrachione (forthcoming), we trained native English-speaking adults to use pitch patterns (or lexical tones) to identify a vocabulary of 6 English pseudosyllables superimposed with 3 pitch patterns (18 words). Successful learning of the vocabulary necessarily entailed learning to use lexical tones in words. We found that learners tended to be more successful, if they had increased musical experience and pitch perception ability. In an accompanying fMRI study, we found successful learners to show greater auditory cortex activation in response to pitch pattern discrimination before training (Wong et al., forthcoming). In the current study, we use the same group of subjects from the previous behavioral and fMRI studies and focus on brain anatomy and spoken language learning, specifically the anatomic characteristics of Heschl's Gyrus (HG) and the ability to use pitch in words.

Recent research has implicated HG to be associated with pitch processing and auditory learning abilities. For example, Gaser and Schlaug (2003) and Schneider et al. (2005) found increased gray matter volume in auditory cortical regions in musicians relative to nonmusicians. Specifically, Schneider et al. (2005) found musicians to have larger lateral HG volume relative to nonmusicians; the size of left HG was especially pronounced, if the individuals had a tendency to rely on F0, as opposed to higher harmonics, to perceive pitch. Neuroanatomic differences were also observed in clinical populations with various auditory-related symptoms, for example, schizophrenia (e.g., Hirayasu et al. 2000), dyslexia (e.g., Leonard et al. 2001; Hugdahl et al. 2003), and congenital deafness (Emmorey et al. 2003; cf. Penhune et al. 2003). In auditory learning, Golestani et al. (2007) found increased left HG white matter to be associated with the successful learning of a rapid (about 40 ms) acoustic cue in nonword contexts (i.e., identifying individual sounds without using them in words).

An important remaining question is whether anatomical differences in HG contribute to the learning of foreign speech sounds in true linguistic contexts (e.g., words). The linguistic and nonlinguistic distinction is an important one. Numerous studies have found that for the same acoustic stimuli, linguistic contexts/functions can modulate cortical responses differently than nonlinguistic contexts (e.g., Gandour et al. [1996]; Wong, Parson, et al. [2004]; see also Gilbert et al. [2001] for a review of perceptual learning studies, including contextual effects). These studies suggest the possibility that success in speech learning lies in the integrity of brain regions that are essential to speech processing alone (i.e., lateral superior temporal region) (e.g., Scott and Wise 2004; Wong, Nusbaum, Small 2004; Liebenthal et al. 2005). This would imply that subtle structural and functional differences in primary auditory regions, such as HG would have little impact on overall speech processing. However, if speech perception, especially during the learning of novel speech sounds, involves a scaffolding process relying on basic acoustic cues, the brain regions sensitive to those cues should also be pertinent. As HG, especially the anterolateral portion, has been implicated in nonlinguistic pitch perception and learning (e.g., Jäncke et al. 2001; Zatorre et al. 2002; Bendor and Wang 2005), the study of linguistically relevant pitch patterns provides a unique opportunity for examining the impact of more primary cortical structures on language processing and learning. Due to the linguistic nature of our lexical learning task, as well as the role of F0 in lexical tone perception, we expect left HG to be associated with success in pitch-to-word learning.

Methods and Materials

Subjects
Subjects were 17 young adult native speakers of American English (ages 18-26 years, mean = 20.65; 10 females), who reported having no audiologic, cognitive, neurologic, or linguistic (word finding, writing, reading, and speech production and comprehension) deficits. All passed a pure-tone audiometric screening bilaterally at 30 dB hearing level for the octave frequencies from 500 to 4000 Hz in a sound attenuated chamber. Subjects were undergraduate students at, or recent graduates

of, Northwestern University. All but 2 subjects were right handed as assessed by the Edinburgh Handedness Inventory (Oldfield 1971) with a score of greater than 40. The remaining subjects, one in each group, had a score of 0 (ambidextrous) and 40 (borderline right handed/ambidextrous). None of the subjects had previous exposure to a tone language at any time in life. All were subjects in our behavioral and fMRI training studies (Wong and Perrachione, forthcoming; Wong et al., forthcoming). Table 1 includes basic demographic information for each subject.

Because musical training has been shown to relate to anatomical variations in auditory areas, the extent of musical training in subjects was assessed by self-report and only subjects who fit our definition of musicians and nonmusicians were included. Eight subjects were amateur musicians (ages 19-26 years, mean = 21.13), defined by at least 6 years of formal private lessons in one instrument starting before the age of 10 (most of the subjects started earlier and had experience with multiple instruments). Nine subjects were nonmusicians (ages 18-25 years, mean = 20.22), defined by no more than 3 years of private lessons in any combination of instruments.

Training Stimuli and Procedures
Subjects were trained to match associate monosyllabic pseudowords with pictures. The key characteristic of these training stimuli was that pitch was used to mark word meaning. Specifically, the training stimuli consisted of 18 English pseudowords with pitch (F0) patterns resembling mandarin tones 1 (level), 2 (rising), and 4 (falling) (the dipping tone (Tone 3), the most complex tone, was excluded to facilitate learning). As shown in Table 2, there were 6 sets of words with minimal pitch contrasts in each set. The 6 base syllables (pesh, dree, ner, vece, nuck, and fute) were originally produced by a native speaker of American English. These syllables were subsequently resynthesized to include variants consisting of the 3 different pitch patterns using the Pitch Synchronous Overlap and Add method implemented in the software Praat (Boersma and Weenink 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, Parson, et al. (2004). 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. Subjects were trained to identify word meanings as depicted by black and white drawings. Word meanings assigned to the stimuli (listed in Table 2) were high-frequency English nouns (Raymer AM, Maher LM,

Table 1

Basic subject demographic information and HG measurements

Subject numbers	Age	Sex	Handedness	L Gray	L White	L Total	R Gray	R White	R Total	L Dup	R Dup
Successful Learners											
S-05-010	21	M	57.14	1963	743	2706	1524	812	2336	D	
S-05-016	20	M	86.67	1605	532	2137	1660	465	2125		D
S-05-052	19	F	100.00	1774	686	2460	1487	437	1924	D	S
S-05-057	20	M	100.00	1579	495	2074	1799	781	2580	S	
S-05-060	20	F	0.00	2163	979	3142	1940	633	2573	D	D
S-05-062	25	F	75.00	1712	539	2251	1318	449	1767		D
S-05-064	19	F	100.00	1372	304	1676	1341	269	1610		
S-05-082	23	F	100.00	1354	548	1902	982	396	1378		
S-05-016	21	F	80.00	1930	703	2633	1372	870	2242	S	
Mean	20.89		77.65	1716.89	614.33	2331.22	1491.44	568.00	2059.44		
Less Successful Learners											
LS-05-055	21	F	88.89	1693	479	2172	1512	345	1857		S
LS-05-061	21	F	85.71	1373	362	1735	1366	395	1761		
LS-05-063	18	M	86.67	764	405	1169	1830	1288	3118		D
LS-05-068	20	M	71.43	1528	551	2079	1388	560	1948	S	D
LS-05-070	18	F	84.62	787	230	1017	1911	676	2587	S	
LS-05-076	26	M	100.00	757	362	1119	737	381	1118	D	D
LS-05-086	18	F	100.00	1404	378	1782	859	467	1326		
LS-05-097	21	M	40.00	1343	712	2055	1036	786	1822	D	D
Mean	20.38		82.17	1206.13	434.88	1641.00	1329.88	612.25	1942.13		

Note: D, Complete Duplication; L, left; R, right; S, Split. L/R "Dup" indicates whether duplication exists.

Greenwald ML, Morris MK, Rothi LJG, Heilman KM 1990, *The Florida Semantics Battery*, unpublished test.). Similar to Curtin et al. (1998), to facilitate learning, the 18 words were divided into 6 groups of 3 stimuli. In a training session, subjects learned to associate a picture with 1 of 18 pseudowords; each word was heard 4 times with its corresponding picture presented, followed by a quiz with feedback on the words they had just learned. At the end of each training session, subjects were presented with the 18 trained words, randomized and repeated 3 times (54 trials total), and were asked to identify each word by selecting the corresponding drawing out of 18 possible choices with no feedback given. The score received from this last word identification test was used to determine whether the training criterion was met. Subjects received 3–4 training sessions per week with no more than one session in a day. The training program was terminated when subjects reached at least 95% accuracy for 2 consecutive sessions or when they failed to improve by at least 5% accuracy for 4 consecutive sessions (the term "asymptotic performance" is defined as the first session in which the successful subjects reached greater than 95% accuracy or when the less successful subjects reached the first of 4 sessions in which they showed no more than 5% improvement). Subjects whose training was terminated because of the former criterion were classified as "successful learners" and those who fell in the latter criterion were classified as "less successful learners." As discussed below, our data analyses were largely based on comparing neuroanatomic differences between these 2 groups of subjects. Further details of the training stimuli and procedures can be found in Wong and Perrachione (forthcoming).

Anatomical Magnetic Resonance Imaging Acquisition and Preprocessing

Subjects in our fMRI study received both functional and anatomical scans from a Siemens Trio 3T scanner before and after training (Wong et al., forthcoming). The T_1-weighted anatomical magnetic resonance (MR) images were acquired sagittally (magnetization prepared rapid gradient echo with a time repetition/time echo of 2100 ms/2.4 ms, flip angle of 8 degrees, time to inversion of 1100 ms, matrix size of 256 × 256, field of view of 22 cm, slice thickness of 1 mm). Only pretraining anatomical scans were used for the present analysis. Similar to other related studies (e.g., Golestani et al. 2007), these images were normalized to a standard stereotaxic space using only linear transformations to avoid warping of pertinent brain structures (Collins et al. 1994). Images were corrected for intensity inhomogeneities using the nu_correct program implemented in the MRI software programs from the Montreal Neurological Institute (Sled et al. 1998).

HG Measurements

T_1-weighted images acquired in the pretraining MR session were used for manually marking HG. The software Display from the Montreal Neurological Institute was used as it allows for the simultaneous viewing of the brain in 3 dimensions, which is crucial for anatomical marking (see Fig. 1). Landmarks for HG delineation were determined based on previously published studies (Rademacher et al. 1993; Penhune et al. 1996; Schneider et al. 2005), and the exact measurement procedures implemented were based on the method of Penhune et al. (1996). The anterior border of HG is defined by the first transverse sulcus, and the posterior border is defined by the first complete Heschl's sulcus. HG may include a sulcus intermedius (SI), which typically does not extend completely lateral medially as in the first complete Heschl's sulcus. In cases of gyral "complete duplications," defined by an SI extending more than half of the anterior HG, as opposed to extending less than half (i.e., a "split" HG), only the most anterior HG was included in the measurements, as cytoarchitectonic studies have shown primary auditory cortex to lie mainly within the first HG (Rademacher et al. 1993). Due to the large variability in the gyral shape, including only the anterior HG in these measurements did not necessarily result in smaller HG volumes. For gray and white tissue classification, a semiautomatic procedure was used as the primary method similar to Penhune et al. (1996). This semiautomatic procedure uses Display for showing the MR signal intensity histograms of the anatomic images. After HG was marked and the total volume measured, the gray/white boundary for the scan was calculated from the histogram by identifying the peak intensity values corresponding to gray and white matter and taking the midpoint. HG volumes were then automatically segmented so that voxels with intensity values below the boundary were labeled as gray matter and those with intensity values above the boundary were labeled as white matter. As an additional (validating) procedure for tissue classification, the software INSECT (Zijdenbos et al. 1998) was used for automatically classifying (segmenting) the tissues within HG into gray and white matter. Volumes of white matter, gray matter, and total HG were recorded for right and left HG of each subject.

Before brain measurements, an individual who did not serve as a rater randomized the brains from all subjects and assigned them with a unique number. This individual also randomly flipped some of the brains so that about half of the brains followed neurologic convention and about half followed radiologic convention. One primary rater (AR) measured HG on all the brains. They were then checked by 2 other individuals (PW and CW) at weekly meetings, and concerns were discussed with AR and consensus developed. One additional rater (AS) marked about 50% (8 out of 17, 4 from each subject group) of the brains; the reliability (Pearson's r), calculated based on total HG volume, was at 0.85 ($P < 0.001$).

Total Cerebral Volume

The software program FreeSurfer (Fischl and Dale 2000) was used to automatically measure total cerebral volume for each subject. As part of its reconstruction process, FreeSurfer removes all nonbrain structures on T_1-weighted scans based on a combination of watershed algorithms and deformable surface models. Total cerebral volume is calculated by counting the number of voxels in the FreeSurfer identified cerebral volumes for each subject.

Results

Based on our definition of successful learning discussed earlier, we found 2 groups of subjects, including 9 "successful learners" and 8 "less successful learners." As discussed in Wong and Perrachione (forthcoming), a 2 × 2 (group × training) repeated measurements analysis of variance (ANOVA) on word identification accuracy at the first session of training and word identification accuracy at the first session of asymptotic performance revealed a main effect of training ($F_{1,15} = 118$, $P < 0.0001$), demonstrating that all subjects improved to a certain extent. For the successful subjects, the mean word identification accuracy at the first session of training and at the first session of asymptotic performance was around 36.63% and 97.12%, respectively, and for the less successful learners, around 27.31% and 63.49%, respectively. We found no significant

Table 2
Subjects were trained on a vocabulary of 18 artificial words

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

Note: 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.
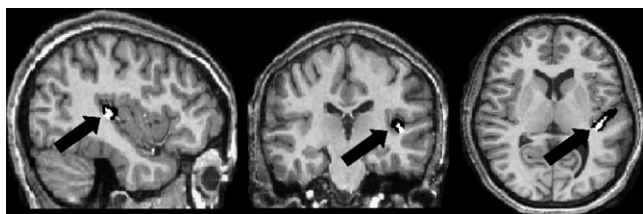
Figure 1. Gray (Black) and white (white) matter within HG of a representative subject shown on sagittal (left), coronal (middle), and axial (right) planes.

difference between the 2 subject groups in the number of sessions it took to reach asymptotic performance; successful subjects as a group took 7.22 (range 2–12) sessions to reach asymptotic performance, whereas less successful subjects took 9.38 (range 5–18) sessions. These 2 groups also did not differ in the type of errors they made (errors could be due to misidentifying the consonants and vowels or the tones of the training stimuli). While the less successful learners made more errors across training in terms of absolute numbers, the percentage of consonant–vowel and tone-only errors made by both groups were the same. Almost all errors made by both groups toward the end of training were tone-only errors, indicating that they both learned the consonants and vowels early on; this also demonstrates that the less successful learners did not have a particular deficit in learning consonants and vowels of the training stimuli.

The 2 groups did not differ in age, height, weight, and handedness scores. They also did not differ in total cerebral volume (successful group: mean = 1 413 480, standard deviation [SD] = 49877.23; less successful group: mean = 1 449 569, SD = 64570.39; $t_{15} = -1.3$, $P = 0.214$).

HG Measurements

Due to the linguistic nature of our lexical learning task, as well as the role of F0 in lexical tone perception, we hypothesized that the left HG would contribute to success in pitch-to-word learning. To assess whether left HG volume is associated with successful learning, left gray and white HG volumes (measured from pretraining scans) from the successful and less successful learner groups were entered into a repeated-measures ANOVA (see Table 1 for individual HG measurements). We found a main effect of tissue ($F_{1,15} = 256.61$, $P < 0.001$), showing gray matter

volume to be larger than white matter volume regardless of subject group. We also found a main effect of group ($F_{1,15} = 9.49$, $P < 0.005$), with the successful learners having larger left total HG volume. A significant group × tissue interaction ($F_{1,15} = 8.02$, $P < 0.02$), driven by increased gray matter in the successful learner group, was also found. A Tukey's honestly significant difference (HSD) post hoc analysis confirmed that gray matter volume was larger in the successful group (see Fig. 2A). There was a trend for white matter volume to be larger in the successful group ($t_{15} = 2.16$, uncorrected $P = 0.047$; $t_{crit} = 2.88$). Data from the automatic tissue classification method confirmed these results. Figure 3 shows representative coronal slides of left HG from one successful learner (Panel A) and one less successful learner (Panel B).

As a control measure for demonstrating that the aforementioned left HG differences were not due to a more general neuroanatomic difference in the auditory cortex, HG volumes from the right hemisphere from each subject group were also entered into a repeated-measures ANOVA. Again, we found a main effect of tissue ($F_{1,15} = 104.37$, $P < 0.001$) but importantly no main effect of group or significant interaction (Fig. 2B).

We also recorded instances of duplications in each subject (Table 1), noting both true duplications when SI extended more than half of HG and splits when SI extended less than half of HG. For the left HG, we found 5 out of 9 and 4 out of 8 instances of duplications regardless of type in the successful and less
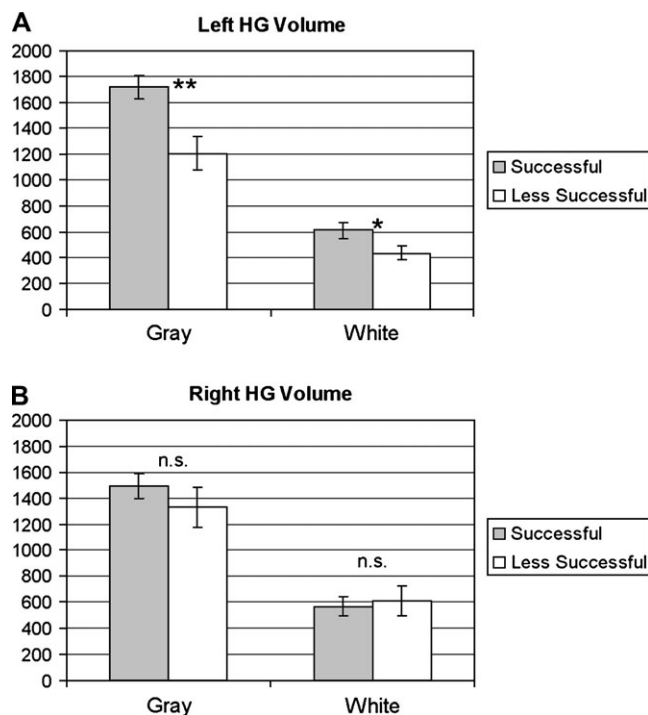
Figure 2. HG Volume in the left (A) and right (B) hemispheres. Error bars indicate standard error of the mean. **$P < 0.007$ and *$P < 0.05$ based on independent-samples t tests.
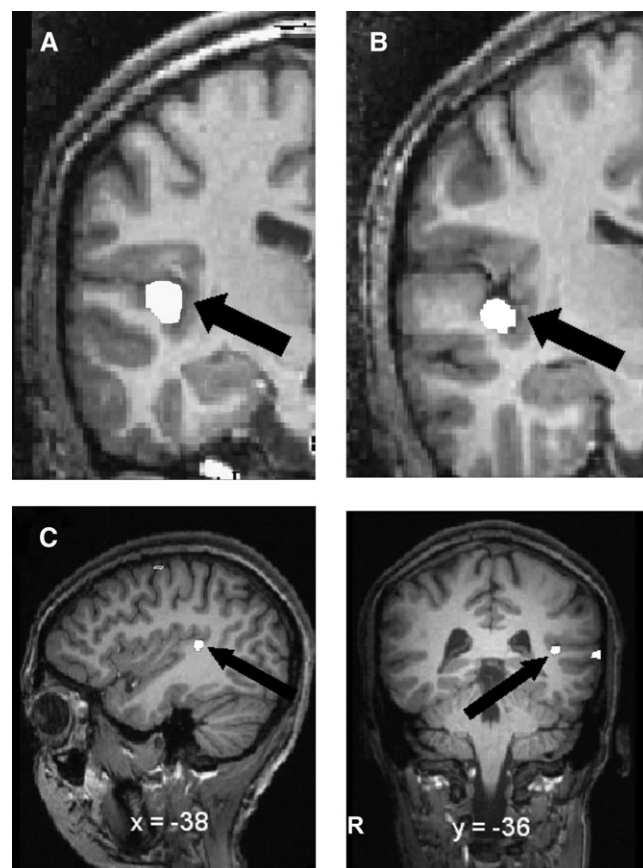
Figure 3. White label shows left HG label from (A) a representative successful learner and (B) a representative less successful learner. Panel (C) shows activation (in white) bordering HG after training in the successful versus less successful learners contrast. Activation (single-voxel $t = 3.3$, $P < 0.001$) is projected onto the brain of one subject for visual clarity (for details see Wong et al., forthcoming).

successful groups, respectively. For the right, we found 4 out of 9 and 5 out of 8 instances, respectively. In other words, frequency of duplications does not appear to be associated with learning success.

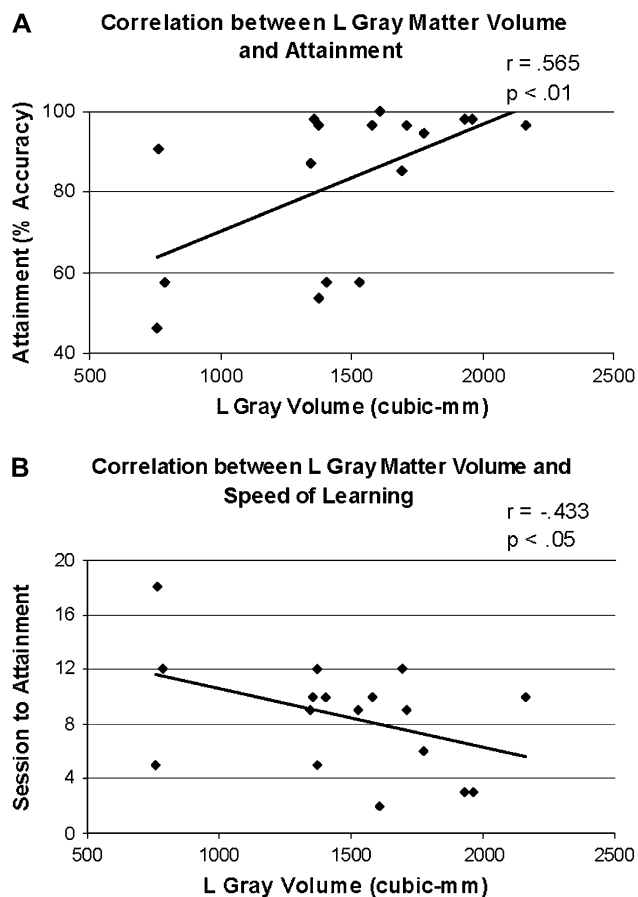### Correlation Analyses: HG Volumes and Behavioral Measures

To examine more specifically the relationships between HG volumes and learning, several correlation analyses were performed. Our training protocol did not provide a specific timeframe for terminating training. Rather, subjects were trained until their individual asymptotic performances were reached. Thus, behavioral measures include their word identification performance at the point of asymptote (henceforth "attainment level"), as well as the number of sessions required to reach that asymptote (henceforth "speed of learning"). It is worth noting that no significant correlation was found between these 2 behavioral measures (i.e., faster or slower learning did not lead to better or worse learning).

We found significant positive correlations between attainment level and left gray (Pearson's $r = 0.565$, $P < 0.01$; Fig. 4A) as well as white (Pearson's $r = 0.547$, $P < 0.02$) matter volume, where larger volumes were associated with higher percent accuracy at the end of training. Moreover, we found a significant negative correlation between speed of learning and left gray matter volume (Pearson's $r = -0.433$, $P < 0.05$; Fig. 4B), indicating that the larger the left gray matter volume, the fewer

A Correlation between L Gray Matter Volume and Attainment

r = .565
p < .01

B Correlation between L Gray Matter Volume and Speed of Learning

r = -.433
p < .05

Figure 4. Correlations between left gray matter volume (regardless of subject group) and (A) attainment level and (B) speed of learning.

sessions it took to reach asymptote. The correlation between speed of learning and left white matter volume was not significant (Pearson's $r = -0.323$, $P = 0.103$). No significant correlations were found between right hemisphere measures and behavioral measures.

Comparison with Previous HG Measurements

Because our procedures for HG measurement were based on the methods of Penhune et al. (1996) and Penhune et al. (2003), we directly compared HG results from the current study with results from the normal-hearing individuals from the other 2 comparable studies. Penhune et al. (1996) reported data from 20 normal-hearing subjects (one of whom lacked gray and white matter segmentation due to technical difficulties), and Penhune et al. (2003) reported data from 10 normal-hearing subjects. Table 3 lists the mean and SD values for HG when data from the 2 previous studies are combined. In a group (previous, successful, and less successful subjects) × hemisphere × tissue repeated-measures ANOVA, we found a main effect of tissue ($F_{1,44} = 151.288$, $P < 0.001$), a significant hemisphere × group interaction ($F_{1,44} = 4.90$, $P = 0.012$), and a marginally significant 3-way interaction ($F_{1,44} = 2.596$, $P = 0.086$). There was no main effect of group. Tukey's HSD post hoc analyses confirmed that previous subjects had significantly larger left white and left total HG volume relative to our less successful subjects only; the difference in left gray matter was marginal ($P = 0.095$). Neither significant differences in the right hemisphere nor significant differences between the previous subjects and our successful subjects were found. Furthermore, we found 3/8 and 5/8 of the less successful subjects showing left gray and white volume measures, respectively, below one SD of the previous data, whereas only 1/9 of the successful subjects showed left white volume below one SD. Taken together, these data suggest a reduction in volume of the less successful group in the left

Table 3
Mean and SD of HG measurements from the normal-hearing subjects reported in Penhune et al. (1996) and Penhune et al. (2003)

	L Gray	L White	L Total	R Gray	R Left	R Total
Mean	1676.90	924.00	2617.77	1443.97	533.76	1977.63
SD	641.98	512.44	1021.16	461.40	283.00	555.10
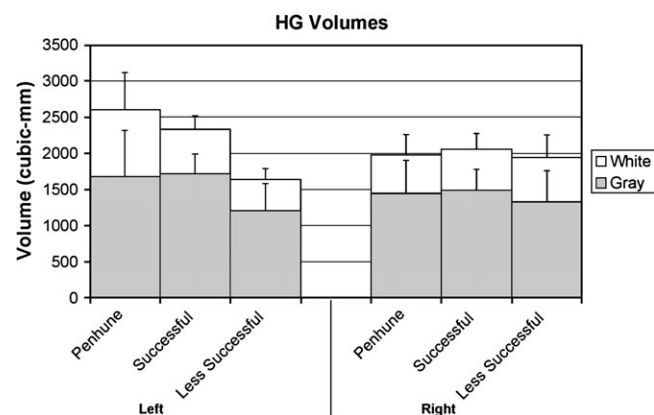
Figure 5. HG volumes found in previous studies (Penhune et al. 1996, 2003) and the current study (both successful and less successful learners). Error bars indicate 1 SD.

hemisphere only, with little difference between the successful group and subjects from the previous 2 studies (see Fig. 5).

Behavioral, Neurophysiologic (Functional), and Neuroanatomic Predictors of Attainment

Because our subjects had all participated in behavioral testing (Wong and Perrachione, forthcoming), fMRI scanning before and after training (Wong et al., forthcoming), and this present neuroanatomic study, we were able to use all 3 factors for predicting attainment. Behaviorally, we have found successful learners (mostly amateur musicians) to score higher in a pre-training, nonlexical, pitch pattern identification test relative to less successful learners. This test involved the identification of the level, rising, or falling pitch patterns embedded in vowels. Neurophysiologically, we found pretraining activation in the auditory cortex to be higher bilaterally in the successful learners. Pretraining activation was calculated based on averaging percent signal change in voxels in the superior temporal gyrus (STG) that exceeded a statistical and 3-dimensional contiguity threshold of $P < 10^{-5}$ and 5 mm³ (based on a Monte Carlo simulation for correcting multiple comparison). In the present study, we found the left HG volume to be greater in the successful learners. As an exploratory measure, we entered all of these 3 factors into a multiple regression analysis simultaneously for predicting attainment level and found an R^2 of 0.609 ($P < 0.01$). Using the backward multiple regression method, the neurophysiologic and neuroanatomic variables were separately removed from the regression model. Removing the neuroanatomic variable from the equation resulted in an R^2 of 0.594 ($P < 0.01$). Removing both the neuroanatomic and neurophysiologic variables resulted in an R^2 of 0.528 ($P < 0.01$). Thus, the behavioral measure alone significantly predicts attainment level and the addition of the other measures augments this model.

Discussion

This is the third of a series of studies examining pretraining behavioral, neurophysiologic (cerebral hemodynamic responses measured by fMRI), and neuroanatomic factors influencing pitch-to-word learning. Behaviorally, we have found pretraining nonlexical pitch pattern (pitch patterns not used in words) identification and musical experience to be associated with learning success (Wong and Perrachione, forthcoming). We also found pretraining neurophysiologic responses in the auditory cortex to be associated with learning (Wong et al., forthcoming). What remained to be established is whether preexisting structural markers have an effect on subsequent learning. In the present study, we found such a neuroanatomic marker for predicting learning success located in HG, which includes the primary and secondary auditory cortical regions important for pitch processing (Zatorre 1988; Patterson et al. 2002; Penagos et al. 2004; Bendor and Wang 2005; for a review see Bendor and Wang 2006). When all of these factors were combined, we found an explanation for a major proportion of variance in attainment level, more so than when only one factor was used. The neuroanatomic marker identified corresponded to the volume (size) of HG, which could be a result of greater thickness, surface area, or both; volume, thickness, and surface area have been found to be correlated with each other (Wiegand et al. 2004; Narr et al. 2005; Hardan et al. 2006).

HG and Speech Processing

In the present study, we found that individuals who successfully incorporated pitch into word contexts showed greater left HG volume (but not right HG volume) relative to those who were less successful. This effect was more pronounced in the gray matter than the white. When all subjects were considered, left HG volume predicted how well and how fast subjects learned. These results suggest the importance of primary cortical structures and adjacent areas even in lexical/linguistic learning. The posterior two-thirds of HG contain the primary auditory cortex which is tonotopically organized (e.g., Merzenich and Brugge 1973; Rademacher et al. 1993), whereas the anterolateral portion contains regions important for pitch processing as evidenced by human lesion studies (e.g., Zatorre 1988), human fMRI studies (e.g., Patterson et al. 2002; Penagos et al. 2004), and animal neurophysiological studies examining the human homologue of this anterolateral region (Bendor and Wang 2005). Thus, by measuring the entire HG, we were not only able to examine the anterolateral portion of HG but also able to consider the primary auditory cortex which provides input to this nonprimary region for making accurate pitch decisions (Bendor and Wang 2006). Furthermore, because the posteromedial and anterolateral portions of HG receive input from the ventral and dorsal medial geniculate body (MBG), respectively, and because these 2 compartments of the MGB encode narrow and broadband auditory signals (Kaas and Hackett 2000), measurements of the entire HG is especially useful for considering a broad range of auditory signals that contain pitch information.

It has been found that speech processing is typically associated with the STG and surrounding areas (auditory association cortex) rather than the HG (primary auditory cortex and adjacent areas; e.g., Liebenthal et al. 2005). For example, it has been found that even though behavioral studies showed F0 to be important in speech perception in mixed-talker compared with single-talker listening situations (Nusbaum and Morin 1992), an fMRI study comparing these 2 types of speech processing found STG, but not HG, activation to differentiate between the 2 conditions (Wong, Nusbaum, Small 2004). In the present study, the size of left HG, especially gray matter, differentiated successful and less successful learner groups. Our results may suggest that the process of learning requires greater perceptual weighting (Nosofsky 1986; Goldstone 1998) of acoustic details processed by HG that more experienced listeners may not need. Behavioral studies of cross-linguistic speech perception suggest that distortion of the speech signals, including the masking of acoustic details, impaired speech perception by nonnative speakers more so than native speakers (e.g., Takata and Nabelek 1990; Garcia Lecumberri and Cooke 2006). Thus, it is possible that increased usage of more primary structures is specific to learning when the listeners are inexperienced with the acoustical signals (such as being nonnative speakers). Interestingly, in our fMRI study in which subjects in the present study participated (Wong et al., forthcoming), we did indeed find a cluster in the vicinity of left HG that activated to a greater degree in the successful learners compared with the less successful learners (Fig. 3C), which could be due to the relatively larger anatomical volume, an increase in physiologic response independent of the anatomical volume, or an increase in both.

It is worth emphasizing that we are not asserting a strict feedfoward model for all auditory processing but are suggesting

that the processing of more basic auditory features may be an important component of lexical learning. Although it may seem obvious that auditory objects cannot be perceived without some level of basic acoustic encoding, there need not be a continuous relationship between basic encoding and higher level processing. For example, it is conceivable that once basic physical encoding is achieved to a certain threshold, further accuracy in encoding does not contribute to better higher level processing. In the context of speech perception, it is possible that higher level processes such as acoustic integration, normalization, and acoustic-phonetic matching (likely supported by STG) would dominate behavioral performance once a minimal/sufficient amount of acoustic information is encoded. For example, a sine wave complex can evoke speech perception (Remez et al. 1981) and activate the STG (Liebenthal et al. 2003) despite the lack of acoustic details. Whereas correlational analyses do not imply causality, our data provides evidence for a continuous relationship between primary auditory regions that contribute to more basic auditory processing and higher level learning.

It is important to point out that when compared with data from 2 previous studies measuring HG volumes in normal subjects (Penhune et al. 1996, 2003), we did not find an enlargement of left HG in the successful subjects but rather a reduction in the less successful subjects. Penhune et al. did not select their subjects based on musicianship, whereas we only included amateur musicians and nonmusicians in the current study. By selecting individuals with less than 3 years of musical training, it is likely that we were selecting individuals who had less musical training than what is typical in university student populations (from which the present and Penhune et al. studies selected subjects). It is perhaps more appropriate to say that our results are more a reflection of less successful, rather than successful, adult spoken language (sound-to-word) learning. Thus, our findings are similar to studies linking neuroanatomic differences (in many cases, anomalies) with various auditory-related symptoms in different clinical populations, such as, schizophrenia (e.g., Hirayasu et al. 2000), dyslexia (e.g., Leonard et al. 2001; Hugdahl et al. 2003), and amusia (Hyde et al. 2006). In the present study, we demonstrate that neuroanatomic differences are associated with differences in adult spoken language (sound-to-word) learning.

The strong left lateralization effect seen in the present study may appear surprising given the consistent evidence for the importance of regions surrounding the right HG in the analysis of pitch information (for a review see Zatorre et al. 2002). However, the right auditory cortex's importance for pitch processing is typically found in nonlinguistic, especially musical, contexts. The present data demonstrate that when pitch information must be integrated into a linguistic task, anatomical features of the left HG are important. Furthermore, prior studies of nonlinguistic pitch processing have emphasized the contribution of right auditory cortex specifically to fine-grained pitch analysis (e.g., Zatorre and Belin 2001), whereas the pitch contours used here span a considerable larger pitch range.

Musicianship
Our results complement studies examining neuroanatomic characteristics of musicians and nonmusicians. These studies showed anatomical differences between musicians and nonmusicians in different auditory cortical regions (e.g., Schneider et al. 2002, 2005; Gaser and Schlaug 2003). With regards to pitch processing and HG, Schneider et al. found increased gray matter volume in both hemispheres in musicians.

Our study complements these studies by showing that decreased HG volume is associated not only with decreased musical and nonlexical pitch perception ability but also with linguistic ability, when lexical tones (pitch) are involved. Our results linking pitch-to-word learning and musicianship are also consistent with studies showing musicians to be better at detecting pitch incongruities in both speech and music (Schön et al. 2004). The results from the present study are also consistent with studies showing native–English-speaking musicians' more accurate encoding of Mandarin tones at the auditory brainstem (Wong et al., 2007), suggesting a common basic precursor (pitch) for higher level processing (speech and music). However, the relatively larger HG found in musicians may be the result of learning-induced plasticity or it could point to a preexisting anatomical variation that helps people excel at musical tasks and thus be more likely to pursue musical training. It is noteworthy that the musicians in our study were amateur musicians, that is, everyday people who happened to have some years of musical training, unlike some of the related studies that included professional musicians (e.g., Schneider et al. 2005).

An important aspect of the current study is that some subjects without musical training also had larger left HG, and some subjects with musical training did not, indicating that perhaps this anatomical difference can be explained by a variety of environmental and genetic influences, including musical training. In other words, it is not always the case that musical training could lead to larger left HG, which in turn could lead to better pitch-to-word learning. HG size, musicianship, and better pitch-to-word learning overlap but not completely. Further research is needed to provide detailed information of the broad results we found.

Acoustic-Phonetic Cues and Second Language Learning
The fact that we found left, rather than right, HG differences between the 2 learner groups could be explained by the general consensus of the left hemisphere being biased for linguistic processing. According to this view, left HG is especially important in the integration of pitch information that is phonetically/lexically relevant. However, our results can also be explained by a more acoustic-based account. Schneider et al. (2005) found that listeners who tend to rely on F0 in pitch perception showed a leftward HG asymmetry (confined to the lateral portion of HG) regardless of musical training, relative to listeners who rely on spectrum frequency who showed a rightward asymmetry. Interestingly, studies of lexical tone perception often show F0 to be the primary acoustic cue. These include behavioral studies with F0 of the stimuli manipulated (e.g., Wong and Diehl 2003) as well as event related potentials studies of the tracking of F0 encoding as revealed by the frequency following response (Krishnan et al. 2005). Although upper harmonics have been shown to contribute to the perception of lexical tones, stimuli employing F0 were still easier to perceive (Stagray et al. 1992). Thus, the reduced left HG volume found in our less successful learners might reflect difficulty processing F0 (missing or not) rather than difficulty processing linguistic stimuli per se.

In a recent study examining relationships between brain anatomy and nonlexical foreign phoneme identification (Golestani et al. 2007), adult native French-speaking subjects were trained to identify Hindi dental and retroflex consonants

that are nondistinctive in French. These consonants were learned in the nonlexical/nonlinguistic context of consonant-/a/. The critical acoustic difference lay in the first 40 ms of the trajectory of the third formant (resonance) frequency. Learners were classified into "faster" and "slower" learner groups depending on the number of training blocks needed to achieve 80% accuracy. The faster learners showed larger left HG white matter volume relative to the slower learners. These results can be attributed to the rapid nature of the acoustic cue, the nonlexical nature of the task, or both of these factors. Unlike the study of Golestani et al., we found left gray matter to be especially relevant, although white matter volume also differentiated successful and less successful subject groups. Our study complements Golestani et al. (2007) by showing that left HG volume is not only associated with rapid temporal processing and nonlexical phonetic/consonant learning but also with the learning of lexically relevant acoustic cues that span the entire syllable. Further studies need to be conducted to examine whether gray/white matter differences contribute to acoustic cue differences (rapid or slow) or types of learning (lexical or nonlexical).

It has been suggested that neural structures in the left hemisphere are biased toward processing linguistic (including lexical) prosodic information, whereas structures in the right hemisphere are biased toward processing paralinguistic prosodic information (e.g., emotion) (for a review see Wong 2002). However, none of the existing studies that we are aware of specifically point to differential roles of primary and primary-like structures in prosodic processing within the 2 hemispheres. If the left auditory association cortex is indeed associated with linguistic processing and learning, it is conceivable that having more accurate information coming from an adjacent primary structure (rather than the same structure on the opposite side of the brain) would be beneficial.

It is worth noting that a recent study showed gray matter density in the left parietal lobe to be positively correlated with second language proficiency but negatively correlated with the age of acquisition (Mechelli et al. 2004). The present study complements the study of Mechelli et al. by connecting a specific acoustic cue (pitch) with a specific neuroanatomic structure (HG), and by considering neuroanatomic contribution even before training has begun.

HG Duplication
Unlike Golestani et al. (2007), who found greater frequency of HG duplication in the faster learners, and Leonard et al. (2001), who found greater frequency of HG duplication in subjects with phonological dyslexia, we did not find HG duplication to be related to learning success. These interstudy differences may reflect the specificity of the connection between HG gray matter volume and learning that requires the use of pitch, or they may simply reflect considerable individual variability in HG duplication.

Conclusion
We found that a combination of behavioral, neurophysiologic, and neuroanatomic factors can explain a majority of the variance seen in pitch-to-word learning. The current study, in particular, points to the importance of neuroanatomic differences, found before training, in predicting learning success. These results not only add to the growing body of literature

showing the direct consequence of structural differences on human behaviors but also point specifically to anatomical contributions to linguistic learning. These findings suggest several new lines of inquiry into the genesis of such structural differences (e.g., genetic and/or environmental factors, including long-term auditory exposure such as musical training), whether different brain structures are tied to different aspects of linguistic learning, differences in the patterns of learning in light of such structural differences, as well as optimal training strategies.

References

Bendor D, Wang X. 2005. The neuronal representation of pitch in primate auditory cortex. Nature. 436:1161-1165.

Bendor D, Wang X. 2006. Cortical representations of pitch in monkeys and humans. Curr Opin Neurobiol. 16:391-399.

Birdsong D. 1999. Introduction: whys and why nots of the critical period hypothesis for second language acquisition. In: Birdsong D, editor. Second language acquisition and the critical period hypothesis. Mahwah (NJ): Lawrence Erlbaum Associates, Inc. p. 1-22.

Boersma P, Weenink D. 2005. Praat: "doing phonetics by computer." (Version 4.3.04). Available at: http://www.fon.hum.uva.nl/pratt/.

Bongaerts T. 1999. Ultimate attainment in L2 pronunciation: the case of very advanced late L2 learners. In: Birdsong D, editor. Second language acquisition and the critical period hypothesis. Mahwah (NJ): Lawrence Erlbaum Associates, Inc. p. 133-160.

Collins DL, Neelin P, Peters TM, Evans AC. 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assisted Tomogr. 18:192-205.

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.

Emmorey K, Allen J, Bruss J, Schenker N, Damasio H. 2003. A morphometric analysis of auditory brain regions in congenitally deaf adults. Proc Natl Acad Sci USA. 100:10049-10054.

Fischl B, Dale AM. 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA. 97:11050-11055.

Fromkin VA. 2000. Linguistics: an introduction to linguistic theory. Oxford: Blackwell.

Gandour J, Potisuk S, Ponglorpisit S, Dechongkit S, Khunadorn F, Boongird P. 1996. Tonal coarticulation in Thai after unilateral brain damage. Brain Lang. 52:505-535.

Garcia Lecumberri ML, Cooke M. 2006. Effect of masker type on native and non-native consonant perception in noise. J Acoust Soc Am. 119:2445-2454.

Gaser C, Schlaug G. 2003. Brain structures differ between musicians and non-musicians. J Neurosci. 23:9240-9245.

Gilbert CD, Sigman M, Crist RE. 2001. The neural basis of perceptual learning. Neuron. 31:681-697.

Goldstone RL. 1998. Perceptual learning. Annu Rev Psychol. 49:585-612.

Golestani N, Molko N, Stanislas D, LeBihan D, Pallier C. 2007. Brain structure predicts the learning of foreign speech sounds. Cereb Cortex. 17:575-582.

Hardan AY, Muddasani S, Vemulapalli M, Keshavan MS, Minshew NJ. 2006. An MRI study of increased cortical thickness in autism. Am J Psychiatry. 163:1290-1292.

Hirayasu Y, McCarley RW, Salisbury DF, Tanaka S, Kwon J, Frumin M, Snyderman D, Yurgelun-Todd D, Kikinis R, Jolesz FA, et al. 2000. Planum temporale and Heschl gyrus volume reduction in schizophrenia—a magnetic resonance imaging study of first-episode patients. Arch Gen Psychiatry. 57:692-699.

Hugdahl K, Heiervang E, Ersland L, Lundervold A, Steinmetz H, Smievoll AI. 2003. Significant relation between MR measures of planum temporale area and dichotic processing of syllables in dyslexic children. Neuropsychologia. 41:666-675.

Hyde KL, Zatorre RJ, Griffiths TD, Lerch JP, Peretz I .2006. Morphometry of the amusic brain: a two-site study. Brain. 129:2562-2570.

Jäncke L, Gaab N, Wüstenberg T, Scheich H, Heinze HJ. 2001. Short-term functional plasticity in the human auditory cortex: an fMRI study. Brain Res Cogn Brain Res. 12:479-485.

Kaas JH, Hackett TA. 2000. Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci USA. 97:11793-11799.

Krishnan A, Xu Y, Gandour J, Cariani P. 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res. 25:161-168.

Leonard CM, Eckert MA, Lombardino LJ, Oakland T, Kranzler J, Mohr CM, King WM, Freeman A. 2001. Anatomical risk factors for phonological dyslexia. Cereb Cortex. 11:148-157.

Liebenthal E, Binder JR, Piorkowski RL, Remez RE. 2003. Short-term reorganization of auditory analysis induced by phonetic experience. J Cogn Neurosci. 15:1-10.

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

Mechelli A, Crinion JT, Noppeney U, O'Doherty J, Ashburner J, Frackowiak RS, Price CJ. 2004. Neurolinguistics: structural plasticity in the bilingual brain. Nature. 431:757.

Merzenich MM, Brugge JF. 1973. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res. 50:275-296.

Miyake A, Friedman N. 1998. Foreign language learning: psycholinguistic studies on training and retention. In: Healy AF, Bourne LEJ, editors. Mahwah (NJ): Lawrence Erlbaum Associates, Inc. p. 339-364.

Narr KL, Bilder RM, Toga AW, Woods RP, Rex DE, Szeszko PR, Robinson D, Sevy S, Gunduz-Bruce H, Wang Y-P, et al. 2005. Mapping cortical thickness and gray matter concentration in first episode schizophrenia. Cereb Cortex. 15:708-719.

Nosofsky R. 1986. Attention, similarity, and the identification-categorization relationship. J Exp Psychol Gen. 115:39-57.

Nusbaum HC, Morin TM. 1992. Paying attention to differences among talkers. In: Tohkura Y, YSaEV-B, editors. Speech perception, production, and linguistic structure. Tokyo (Japan): Ohmasha Publishing. p. 113-134.

Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9:97-113.

Patterson RUS, Johnsrude IS, Griffiths TD. 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron. 36:767-776.

Penagos HM, Melcher JR, Oxenham AJ. 2004. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci. 24:6810-6815.

Penhune VB, Cismaru R, Dorsaint-Pierre R, Petitto LA, Zatorre RJ. 2003. The morphometry of auditory cortex in the congenitally deaf measured using MRI. Neuroimage. 20:1215-1225.

Penhune VB, Zatorre RJ, MacDonald JD, Evans AC. 1996. Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. Cereb Cortex. 6:661-672.

Rademacher J, Caviness VS Jr, Steinmetz H, Galaburda AM. 1993. Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cereb Cortex. 3:313-329.

Remez RE, Rubin PE, Pisoni DB, Carrell TD. 1981. Speech perception without traditional speech cues. Science. 212:947-950.

Schneider P, Scherg M, Dosch HG, Specht HJ, Gutschalk A, Rupp A. 2002. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nat Neurosci. 5:688-694.

Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht H, Dosch H, 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.

Schön D, Magne C, Besson M. 2004. The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology. 41:341-349.

Scott SK, Wise RJS. 2004. The functional neuroanaotmy of prelexical processing in speech perception. Cognition. 92:13-45.

Shih C-L. 1988. Tone and intonation in Mandarin. In: Working papers of the Cornell phonetics laboratory. CLC Pulications. 3:83-109.

Sled JG, Zijdenbos AP, Evans AC. 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 17:87-97.

Stagray JR, Downs DD, Sommers RK. 1992. Contributions of the fundamental, resolved harmonics, and unresolved harmonics in tone-phoneme identification. J Speech Hear Res. 32:1406-1409.

Takata Y, Nabelek AK. 1990. English consonant recognition in noise and in reverberation by Japanese and American listeners. J Acoust Soc Am. 88:663-666.

Wiegand LC, Warfield SK, Levitt JJ, Hirayasu Y, Salisbury DF, Heckers S, Dickey CC, Kikinis R, Jolesz FA, McCarley RW, et al. 2004. Prefrontal cortical thickness in first-episode psychosis: a magnetic resonance imaging study. Biol Psychiatry. 55:131-140.

Wong PCM. 2002. Hemispheric specialization of linguistic pitch patterns. Brain Res Bull. 59:83-95.

Wong PCM, Diehl RL. 2003. Perceptual normalization of inter- and intra-talker variation in Cantonese level tones. J Speech Lang Hear Res. 46:413-421.

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

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

Wong PCM, Perrachione TK. Forthcoming 2007. Learning pitch patterns in lexical identification by native English-speaking adults. Appl Psycholinguist.

Wong PCM, Perrachione TK, Parrish TB. Forthcoming. Neural characteristics of successful and less successful speech and word learning in adults. Hum Brain Mapp.

Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci. 10:420-422.

Zatorre RJ. 1988. Pitch perception of complex tones and human temporal-lobe function. J Acoust Soc Am. 84:566-572.

Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. Cereb Cortex. 11:946-953.

Zatorre RJ, Belin P, Penhune VB. 2002. Structure and function of auditory cortex: music and speech. Trends Cogn Sci. 6:37-46.

Zijdenbos A, Forghani R, Evans A. 1998. Automatic quantification of MS lesions in 3D MRI brain data sets: validation of INSECT. In: Medical Image Computing and Computer-Assisted Intervention Conference; 1998 Oct 11-13; MA. Berlin: Springer. p. 439-448.