OXFORD

ORIGINAL ARTICLE

Repetition Suppression in the Left Inferior Frontal Gyrus Predicts Tone Learning Performance

Salomi S. Asaridou[1,2], Atsuko Takashima[2,3], Dan Dediu[1,2], Peter Hagoort[1,2], and James M. McQueen[1,2,3]

[1]Neurobiology of Language, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, [2]Donders Institute for Brain, Cognition and Behaviour, and [3]Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands

Address correspondence to Salomi S. Asaridou, Department of Neurology, University of California, Irvine, Biological Sciences III, Room 3230, Irvine, CA 92697, USA. Email: sasarido@uci.edu

Abstract

Do individuals differ in how efficiently they process non-native sounds? To what extent do these differences relate to individual variability in sound-learning aptitude? We addressed these questions by assessing the sound-learning abilities of Dutch native speakers as they were trained on non-native tone contrasts. We used fMRI repetition suppression to the non-native tones to measure participants' neuronal processing efficiency before and after training. Although all participants improved in tone identification with training, there was large individual variability in learning performance. A repetition suppression effect to tone was found in the bilateral inferior frontal gyri (IFGs) before training. No whole-brain effect was found after training; a region-of-interest analysis, however, showed that, after training, repetition suppression to tone in the left IFG correlated positively with learning. That is, individuals who were better in learning the non-native tones showed larger repetition suppression in this area. Crucially, this was true even before training. These findings add to existing evidence that the left IFG plays an important role in sound learning and indicate that individual differences in learning aptitude stem from differences in the neuronal efficiency with which non-native sounds are processed.

Key words: fMRI adaptation, individual variation, inferior frontal gyrus, learning, repetition suppression

Introduction

Learning a second language can be a demanding enterprise, especially when it comes to learning a non-native phonology. Individuals vary greatly in their ability to learn to perceive and produce non-native speech sounds (Golestani and Zatorre 2009; Chandrasekaran et al. 2010; Hanulíková et al. 2012). Although several experience-related factors such as age at which the non-native phonology is acquired (Flege et al. 1999), the amount of exposure to the non-native language (Flege et al. 1997), the overlap between native and non-native phonology (Best et al. 2001), or the amount of music education received (Wong and Perrachione 2007) might all be contributing to this variability, they cannot fully account for it. What is then driving these individual differences?

It has been proposed that individual differences in language-learning aptitude arise, to some extent, as a consequence of individual differences in the functional properties of underlying brain mechanisms (Zatorre 2013). These neuronal predispositions interact with language experience, making some individuals more successful learners than others. A number of training studies have shown that successful learners of non-native speech contrasts process sounds differently compared to less successful learners (Wang et al. 2003; Golestani and Zatorre 2004; Wong et al. 2007; Ventura-Campos et al. 2013). These processing differences can even sometimes be observed before the commencement of training (Wong et al. 2007). The fact that learning attainment correlates with the post-training neuronal activation associated with the non-native sounds is interpreted as showing that

training increases processing efficiency in successful learners (Golestani and Zatorre 2004). The more specific question that then arises is: Are successful learners processing these sounds more efficiently?

fMRI adaptation is a good measure of neuronal processing efficiency. fMRI adaptation or repetition suppression refers to the reduction observed in the BOLD response when a stimulus or stimulus properties are repeatedly presented (Grill-Spector et al. 2006). Although the neurophysiological mechanisms underlying adaptation phenomena are still not fully understood (Segaert et al. 2013), repetition suppression can be interpreted as a neuronal marker of increased processing efficiency (Grill-Spector et al. 2006; Race et al. 2009), such that the more efficient the processing of a stimulus, the greater the BOLD suppression.

A training study by Chandrasekaran et al. (2012) provided evidence that repetition suppression to non-native sounds reflects individual differences in the efficiency with which individuals process non-native sound information. fMRI adaptation was measured in the inferior colliculus (IC), a region in the brainstem which encodes sound frequency (Yan et al. 2005), before participants received training in non-native Mandarin tones. Individuals who showed repetition suppression to tonal contours in the IC prior to training initiation were subsequently better learners of tones (Chandrasekaran et al. 2012). Although the implications of these findings are very interesting, the study focused exclusively on the IC. Pitch processing, however, involves a number of cortical and subcortical areas along the auditory pathway, including the thalamus, the primary and secondary auditory cortices (Javad et al. 2014), as well as frontal areas (Nan and Friederici 2013). Moreover, auditory learning and tuning of subcortical areas relies heavily on their feedback connections to cortical resources (Bajo et al. 2010). Processing efficiency might therefore be reflected in the activity of a specific node in the pathway, or in the orchestration of multiple nodes, that is, efficiency might be instantiated in a stronger connection between the nodes along the pitch processing pathway.

In the current study, therefore, we investigated adaptation effects across the entire pitch processing pathway and asked how they relate to individual variation in tone learning. Using a learning paradigm, we trained Dutch native speakers in non-native pitch contours, modeled after Mandarin tones, over the course of 5 separate sessions. Participants' repetition suppression to the non-native tones was measured at 2 different time points, before and after training. Standard Dutch (the official language taught in school and used in public discourse) does not use tones at the lexical level. Given this, and the results of previous studies using tone training in English speakers (Wong et al. 2007; Chandrasekaran et al. 2012), we anticipated a large individual variation in learning performance.

The purpose of the study was 3-fold. First, we were interested in which area(s) along the pitch processing pathway show repetition suppression to tone, when other acoustic properties (voice and phonemes) vary randomly. Given the hierarchical nature of pitch processing (Javad et al. 2014), we expected that the regions involved in abstracting tonal pitch contours over and above other varying acoustic information would include the bilateral superior temporal gyri (STGs)/sulci and the inferior frontal gyri (IFGs; Wang et al. 2003; Wong et al. 2007; Nan and Friederici 2013). These areas act, in concert, with superior temporal areas being involved in the sensory processing of varying pitch (Javad et al. 2014), whereas the inferior frontal areas, especially in the left hemisphere, being involved in higher-order, decision-making aspects of pitch processing (Nan and Friederici 2013).

The second purpose of the study was to test whether repetition suppression to tone is associated with differences in tone learning success. In other words, we wished to assess the hypothesis that successful learners should process tones more efficiently and therefore show larger repetition suppression when a tone is repeated compared to less successful learners, especially after training. Previous language-learning studies have demonstrated that activation in the left IFG after training is associated with successful tone (Wong et al. 2007) and phonetic learning (Golestani and Zatorre 2004; Ventura-Campos et al. 2013).

Although there is, to the best of our knowledge, no prior fMRI adaptation study correlating repetition suppression to tone with tone learning performance, a study looking at non-native phonetic category learning found a positive correlation between repetition suppression to non-native phonemes and performance in the left IFG (Myers and Swan 2012). Given the existing literature and the evidence for the involvement of the left IFG in tone perception, we expected that repetition suppression in this area (at least after training) would be associated with tone learning performance.

The third purpose of the study was to assess how repetition suppression effects are influenced by changes introduced by learning in the connectivity patterns between pitch processing areas. That is, we were interested in the dynamic changes in feedback and feed-forward connections along the pitch processing pathway that could mediate perceptual learning (Ahissar et al. 2009; Bajo et al. 2010). For that reason, we performed functional connectivity analyses looking at cortical and subcortical areas (i.e., the IC and the auditory thalamus). These areas are involved in pitch processing through afferent and efferent connections to the cortex (Javad et al. 2014).

Materials and Methods

Participants

Forty young adults (15 males, mean age = 22.62, SD = 3.16) participated in the study. They were native speakers of Dutch, and recruited from the Radboud University and Max Planck Institute for Psycholinguistics databases in Nijmegen, the Netherlands. Left-handed participants as well as participants with neurological, speech, or language disorders were excluded from the sample. Participants were all screened for hearing with an Oscilla USB-330 audiometer (Inmedico©, Denmark) using the random automatic hearing test at 20 dB in 11 frequencies ranging from 125 Hz to 8 kHz in both ears. All were able to detect frequencies ranging from 250 Hz to 4 kHz at intensity higher than 30 dB in both of the 2 ears. None of the participants had had experience with a tone language and/or with the tonal dialect spoken in the Dutch province of Limburg. All participants gave written informed consent prior to the experiment (local ethics committee CMO region Arnhem–Nijmegen, the Netherlands) and were compensated with 60 euro or 6 course points.

Stimuli

In the training study, there were 24 Dutch-Chinese hybrid monosyllabic nonsense words. These hybrids (hereafter "Dutchinese") were Dutch in the sense that they were pseudowords with phonemes which followed Dutch phonotactic rules, and Chinese in that Mandarin tone contours were superimposed on the syllables. By using hybrid stimuli, we made sure that participants did not have to learn anything about Mandarin segmental

phonology while at the same time we could create minimal quadruplets differing only in pitch contour with all the other variables (e.g., word duration, intensity, vowel length, production rate etc.) kept constant. The idea was to make the pitch contour the only acoustic information available in the stimuli for participants to dissociate words within a quadruplet.

Seventeen pseudowords with a consonant–vowel–consonant (CVC) structure were created, 6 of which were used for the training paradigm (Table 1). The remaining 11 words were used in the tone discrimination and tone identification tasks. We recorded 8 Dutch native speakers (4 men and 4 women) reading aloud the list of pseudowords at a pace and pitch of their preference. Similarly, we recorded 8 native speakers of Chinese (4 men and 4 women) uttering the word "mi" on 4 citation style Mandarin tones: High level Tone 1 (T1), low rising Tone 2 (T2), low dipping Tone 3 (T3), and high falling Tone 4 (T4). Recordings were made in a soundproof booth using the Adobe Audition software at a 44 100-Hz sampling rate. The hybrid stimuli were then created automatically by superimposing the Mandarin pitch contours on the Dutch utterances using the Functional Data Analysis (FDA) method for speech analysis and re-synthesis (http://lands.let.ru.nl/FDA/index.htm, Last accessed 26/05/2015; Gubian 2011).

Stimulus Ratings

We conducted a rating study to identify the Dutchinese hybrid tokens in which native Mandarin speakers could most correctly and reliably identify the intended Mandarin tone. Twenty-nine Mandarin Chinese speakers were asked to recognize the tone in the hybrid word and rate its naturalness. We then selected the hybrid words spoken by 4 different hybrid Dutch-Mandarin pairs of speakers (hereafter 4 "Dutchinese" speakers) who were most accurately identified and had received the highest naturalness rating.

Dutchinese Training

The training was designed based on Chandrasekaran et al. (2010), adapted to 5 sessions of training. Neurobehavioral Systems Presentation software (www.neurobs.com, Last accessed 26/05/2015) was used for stimulus presentation and response recording. The

Table 1 IPA transcriptions of the hybrid words used in the experiment

Task	Dutch CVC	IPA transcription
Dutchinese training	baaf[a]	[ba·f]
	din	[d··n]
	jor	[j·r]
	moep[a]	[mup]
	nuuk	[nyk]
	wum[a]	[ʋ·m]
Tone discrimination	dul	[d·l]
	goel	[χul]
	luug	[lyχ]
	rof	[r·f]
	tar	[tɑr]
	ziem	[zim]
Tone identification	beem	[be·m]
	nal	[nɑl]
	seek	[se·k]
	wot	[ʋ·t]
	zun	[z·n]

[a]Words used in the fMRI adaptation task.

participants' task was to learn 24 word–picture associations over the course of the 5 training sessions. Each session would start with the training part followed by the testing part. During training, participants were presented with one of the colored pictures of everyday items (from the Snodgrass and Vanderwart set; Rossion and Pourtois 2004) on a computer screen and heard their Dutchinese names from a pair of headphones. To facilitate learning, the presentation was blocked per CVC (6 CVC = 6 blocks) and sub-blocked per Dutchinese speaker. All the items were presented twice for each speaker sub-block with a total of 32 stimuli pairs per block (1 CVC × 4 tones × 2 repeats × 4 speakers) and a total of 192 training trials. Participants were thus trained in each minimal quadruplet for each block. To boost their memory with an emphasis on the tonal differences as the discriminating factor between phonemically identical words, after each block they received a mini-quiz consisting of 16 trials (1 CVC × 4 tones × 4 speakers) in which they were presented with the 4 pictures on the screen, heard one word at a time, and had to click with the mouse on the picture that corresponded to the word. Upon clicking a picture, they would hear the word again and get visual feedback on their response (either the printed word "correct" if they were right or the correct picture if they were wrong; Fig. 1). The training data were not analyzed.

During the testing part, participants were presented with one word at a time and had to click on the corresponding picture from the whole set of 24 presented on the screen. The total number of trials was 96 (6 CVC × 4 tones × 4 speakers) and no feedback was provided during this part. In the final session (Session 6), participants performed a generalization test, which was identical to the regular testing part with the exception that the Dutchinese speakers uttering the words were new (i.e., the other 4 hybrid speakers). Participants' response accuracy was recorded (percentage of correct picture–word matches). As in Chandrasekaran et al. (2010), we took accuracy in the final generalization test as participants' final learning score. Each training–testing session lasted around 30 min in total.

Tone Discrimination and Identification Tasks

Participants completed 2 tone perception tasks prior to training initiation and after training completion [designed after Chandrasekaran et al. (2010)]. The purpose of these tasks was to ensure that the lexical training indeed trained participants in the non-native tone contrasts instead of just tapping into simple associative learning abilities. In the tone discrimination task, participants listened to minimal pairs of Dutchinese words and had to report whether or not the words differed in tone. The pairs were CVC words chosen from 6 minimal tone quadruplets and were different from the ones participants were trained on [see Table 1 for the International Phonetic Alphabet (IPA) transcription]. All the words were uttered by the same female Dutchinese hybrid speaker, so that the only acoustic difference between a pair was the pitch contour. The words were presented using in-house software through headphones with 500 ms inter-stimulus interval, and participants were instructed to press 1 of 2 buttons on a button box as soon as they had made their same–different decision. The task included 8 practice trials with feedback in the beginning and 144 test trials including all possible combinations of tones. Button and trial orders were counterbalanced across participants. Response accuracy was recorded.

In the tone identification task, participants listened to single Dutchinese words and had to indicate the direction of the pitch contour in the word. There were 3 possible directions: upwards (indicated by an upward pointing arrow), downwards (indicated

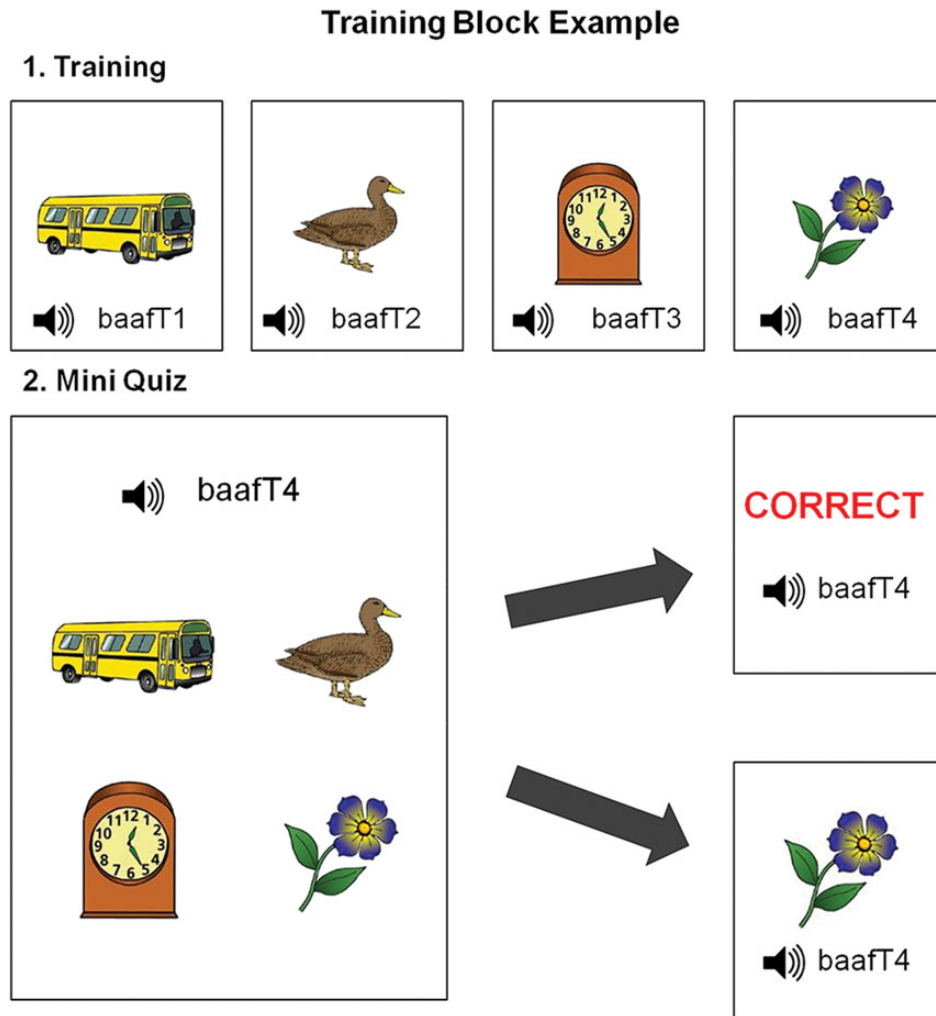## Training Block Example

1. Training

2. Mini Quiz

Figure 1. Example of a Dutchinese training block in which the participant was asked to learn the association between words in the minimal quadruplet baafT1, baafT2, baafT3, and baafT4 and their matching pictures.

by a downward pointing arrow), and flat (indicated by a horizontal flat arrow). The words used in this task were different from the ones used in the discrimination and training tasks (Table 1), and consisted of 5 CVC words uttered by a female and 2 male speakers. After a fixation cross, the word was presented through headphones together with the 3 arrows were presented on the screen. Participants were instructed to listen carefully and click the button corresponding to the correct arrow. The task included 18 practice trials with feedback in the beginning of the test and 135 test trials. Response accuracy was recorded.

Control Tasks

Since learning abilities are influenced by general intelligence and memory abilities, we administered 2 control tasks to assess these abilities in our sample. We used Raven's Advanced Progressive Matrices Test (1998 Edition, set II) to assess non-verbal general intelligence, and the Backward Digit Span (DS) subtest adapted from the Dutch version of the Wechsler Adult Intelligence Scale (WAIS) to assess working memory. Participants were also asked to fill out a post-study questionnaire about their language and music background as well as their motivation and the learning strategies they used during the training.

fMRI Adaptation Task

During the fMRI adaptation task, participants were lying in the scanner and were presented with Dutchinese words through in-ear MR compatible earbuds (Sensimetrics S14 system). The presented words were a subset of the Dutchinese words they were trained on ("baaf", "moep," and "wum") uttered by 2 female speakers. To reduce any influence of expectation, prediction, and attention on our fMRI adaptation effects (Segaert et al. 2013), we used a slow event-related instead of a block design while participants were asked to perform a task that was orthogonal to our measure of interest. As in Chandrasekaran et al. (2012), they performed an intensity judgment in each trial, that is, they reported whether the intensity of the presented word had changed or remained the same compared with the previous one. The task ensured that participants were attending to the words during the experiment.

Each trial began with a white fixation cross presented for a jittered interval of 3–7 s after which the fixation cross turned blue for 1 s followed by the word presentation. After another jittered interval of 3–7 s, participants were presented with the 2 response options on the screen ("same–different") and had to press the corresponding button with their right index or middle finger (Fig. 2B).
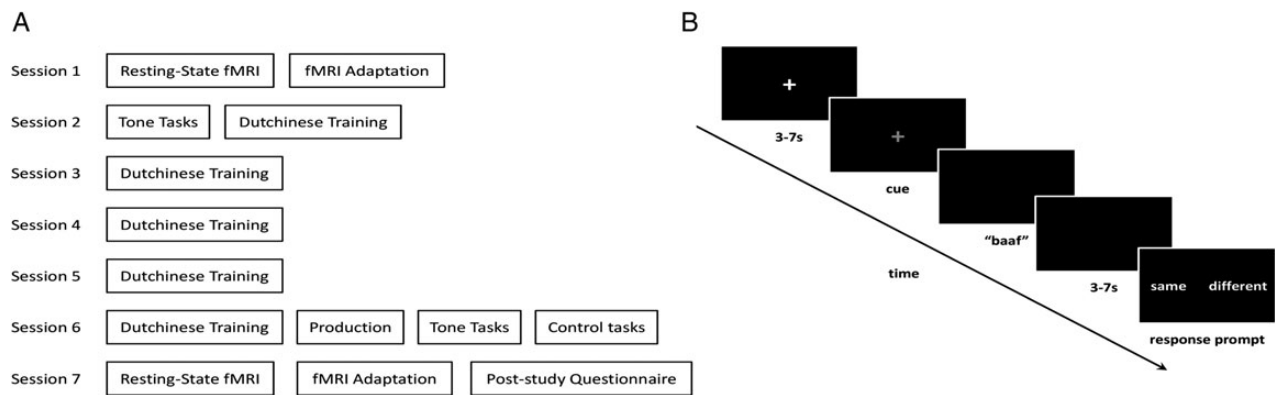
Figure 2. (A) Outline of the experimental procedure. (B) Example of a trial in the fMRI adaptation experiment.

The intensity changed by 65 ± 10 dB in 7% of catch trials. At the same time, however, the tone in the presented words was repeated in 50% of the trials while the other acoustic dimensions varied pseudorandomly. The stimulus list was created using the MIX algorithm (http://www.mrc-cbu.cam.ac.uk/people/maarten-van-casteren/mixandmatch/, Last accessed 26/05/2015). The total number of trials was 364 (including 20 null event trials in which no stimulus was presented) and the task lasted around 35 min. The sound amplitude was adjusted to the participants' comfort level over the scanner noise prior to task initiation. The fMRI Adaptation task took place twice, once before the Dutchinese training, on Session 1 (pre-training), and again after completing the Dutchinese training, on Session 7 (post-training). A post-scanning questionnaire was administered after the last fMRI session to identify participants who could have become aware of the tone manipulation.

Image Acquisition

MRI data were acquired on a Siemens 3T MAGNETOM Trio Tim MR system (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil. We used multiecho planar imaging (EPI) for the functional T_2^*-weighted images where a single excitation was followed by multiple acquisition times. We opted for this type of sequence since it reduces artifacts caused by signal drop-out, which usually affect the inferior frontal and temporal areas we were interested in (Poser et al. 2006). We used a repetition time (TR) of 2.25 s with 4 acquisition times (TEs) at 17 ms (TE1), 26 ms (TE2), 35 ms (TE3), and 45 ms (TE4), with 90° flip angle, accelerated with GRAPPA parallel imaging (accelerating factor 4). We acquired 35 axial slices per volume in an ascending order, with 3 mm slice thickness, 224 mm field of view (FOV), 0.51 mm slice gap, and matrix size 64×64. This allowed us to acquire almost the whole brain, with the exception that the cerebellum was not scanned in most participants. We also acquired a high-resolution T_1-weighted anatomical image using a magnetization-prepared rapid gradient echo sequence with the following parameters: TR: 2.3 s, TE 3.03 ms, 8° flip angle, 192 slices, $1.0 \times 1.0 \times 1.0$ mm^3 voxel size, 256 mm FOV, and matrix size 256×256, accelerated with GRAPPA parallel imaging (accelerating factor 2).

Procedure

The experiment consisted of 7 separate sessions that lasted a total of 7 hours (Fig. 2A). On Session 1, participants performed the pre-training fMRI adaptation task in the scanner. Resting-state fMRI and DTI scans were also collected during that session, but will not be discussed here. On Session 2, participants came to the behavioral laboratory and performed the tone perception tasks (discrimination and identification) as well as the first Dutchinese training–testing task. On Sessions 3, 4, and 5, participants performed the Dutchinese training–testing task only. On Session 6, they performed the last Dutchinese training and generalization testing, followed by the tone perception and the general control tasks (Raven and Backward Digit Span). The training sessions took place on separate days with no more than 3 days between sessions. On Session 7, participants came to the MRI laboratory for the post-training fMRI adaptation task. Resting-state fMRI and an anatomical scan were also recorded. The time between Sessions 6 and 7 was not more than 3 days. Participants were asked to fill out the post-study questionnaire upon completion of the study.

Behavioral Analyses

The behavioral analyses were carried out using the IBM SPSS 19 statistical package. For the Dutchinese training task, participants' response accuracy in matching the Dutchinese words to their corresponding pictures was analyzed using repeated-measures ANOVA, with session (×5) as a factor and percentage correct as the dependent measure. All post hoc pairwise comparisons were Bonferroni-corrected. The tone discrimination and identification tasks were analyzed using paired-sample T-tests to compare mean response accuracy (percentage correct) before and after training. We also performed pairwise correlations between the final learning score and the tone perception tasks as well as the general control tasks, music training duration, and motivation.

fMRI Analyses

Preprocessing
One participant was excluded from the imaging analyses because a brain anomaly was found (as assessed by a radiologist). Seven participants were further excluded from the fMRI analyses (3 did not fulfill the inclusion criteria, being either left-handed or had neurological/speech/language disorders, and 4 due to technical problems).

Since we used a multiecho sequence (i.e., acquired 4 echoes per TR), we combined the echoes before applying any preprocessing

by following the echo-weighting procedure described in Poser et al. (2006). First, all the first echo volumes acquired were realigned to the first volume of the first echo. All the volumes of all the remaining echoes were subsequently realigned to the first echo and resliced. Next, the first 30 acquired volumes were smoothed with a 3-mm Gaussian kernel and used to calculate the optimal echo-weighting parameters (optimal contrast to noise ratio) for combining the echoes. The weighting parameters were subsequently applied to combine the echoes in all the remaining volumes. A mean functional image and a text file with the realignment parameters were created as well.

The next preprocessing steps were performed using SPM8 (www.fil.ion.ucl.ac.uk, Last accessed 26/05/2015). The first 5 functional volumes for each participant were discarded from further analysis to remove non-equilibrium effects of magnetization. The mean functional image was co-registered to the participant's T_1-weighted anatomical image using normalized mutual information, and the registration parameters were subsequently applied to all the functional images. The anatomical image was segmented into gray matter, white matter, and cerebrospinal fluid, and the normalization parameters from the segmentation procedure implemented in SPM8 were used for normalizing and transforming the structural and functional images to the standard Montreal Neurological Institute (MNI) space (2 × 2 × 2 mm voxel size). Finally, all functional images were convolved with a Gaussian smoothing kernel of full-width 8 mm at half maximum.

fMRI Adaptation Statistics

The statistical analysis was performed using a standard general linear model (GLM) approach in SPM8. The model included 4 experimental factors: tone, voice, CVC, and session in a 2 (tone repeat, tone change) × 2 (voice repeat, voice change) × 2 (CVC repeat, CVC change) × 2 (pre-training session, post-training session) factorial design, which resulted in 8 different conditions per session (Table 2). Each trial was defined by the trial preceding it; that is, a trial was classified as belonging, for example, to the tone repeat and voice repeat and CVC repeat (TreVreCre) condition if it shared the same tone, voice, and CVC with the previous trial and the tone change and voice change and CVC change (TcVcCc) condition if all 3 features changed. The first trial, null event trials, and amplitude change trials were modeled in separate regressors. Events were modeled after a stick function (0 s duration), time-locked to word onset, and convolved with the

canonical hemodynamic response function. The 6 realignment parameters, their derivatives, and the squared derivatives (in total 24) were also included in the models as regressors of no interest. Data were high-pass filtered at 128 Hz cutoff and the GLM was estimated using the Restricted Maximum Likelihood (ReML) algorithm in SPM8. T-contrast images for the 16 experimental conditions versus implicit baseline were estimated for each participant and were subsequently entered in a second-level random-effects analysis with random subject effects for population inferences. Since we were interested in adaptation to tone, over and above voice, and consonantal information, we estimated the repetition suppression effect to tone with the following contrast: (TreVreCre + TreVcCre + TreVreCc + TreVcCc) − (TcVreCre + TcVcCre + TcVreCc + TcVcCc) masked exclusively by the repetition suppression effect to voice [(TreVreCre + TreVreCc + TcVreCre + TcVreCc) − (TreVcCre + TreVcCc + TcVcCre + TcVcCc)] and CVC [(TreVreCre + TreVcCre + TcVreCre + TcVcCre) − (TreVreCc + TreVcCc + TcVreCc + TcVcCc)] (mask uncorrected at $P = 0.05$). Results were initially voxel-wise thresholded at $P = 0.001$ (uncorrected) and subsequently, suprathreshold cluster extent was tested using random field methods (Hayasaka and Nichols 2003), corrected for multiple comparison at FWE $P = 0.05$.

Region-of-Interest Analysis

We performed a region-of-interest (ROI) analysis on anatomically predefined regions along the auditory processing pathway. The ROI analysis aimed to increase sensitivity in detecting repetition suppression effects in brain areas that have been reported to process acoustic changes. The ROIs included Heschl's gyri (HGs), STGs, and IFGs bilaterally (Schönwiesner et al. 2007). We also chose to include the left IC based on the findings by Chandrasekaran et al. (2012), and the medial geniculate thalamic nuclei (MGB) since they relay acoustic information from the IC to cortical auditory areas (Javad et al. 2014). The cortical ROIs were defined using the AAL template (Tzourio-Mazoyer et al. 2002) provided by the WFU PickAtlas toolbox (Maldjian et al. 2003) and transformed into the MNI space in MarsBaR (http://marsbar.sourceforge.net/, Last accessed 26/05/2015). The subcortical ROIs (Fig. 3) were defined as spheres using the MNI coordinates reported by Mühlau et al. (2006) (5 mm radius sphere around −6, −33, −11 for the left IC and 8 mm radius sphere around ±17, −24, −2 for the thalamus)

Table 2 fMRI adaptation experimental conditions

Conditions	Factors		
	Tone	Voice	CVC
TreVreCre	Repeat	Repeat	Repeat
TreVcCre	Repeat	Change	Repeat
TreVreCc	Repeat	Repeat	Change
TreVcCc	Repeat	Change	Change
TcVreCre	Change	Repeat	Repeat
TcVcCre	Change	Change	Repeat
TcVreCc	Change	Repeat	Change
TcVcCc	Change	Change	Change
Null events			
TNI			

T, tone; V, voice; C, CVC; re, repeat; c, change; Null events, trials with 20 s of silence and black screen; trials of no interest (TNI), include the first trial and the trials with amplitude change.
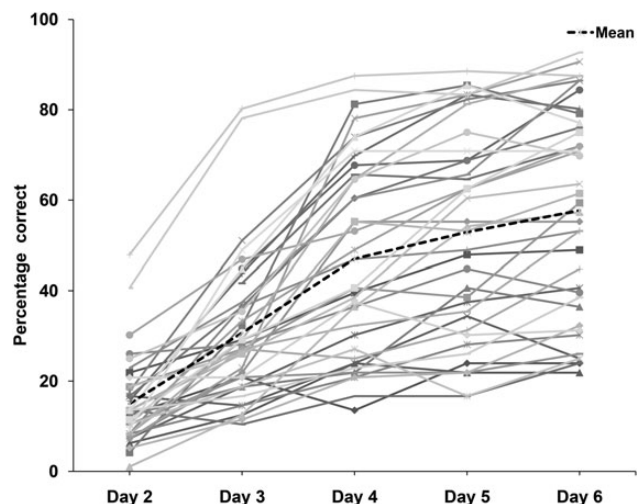

Figure 3. Individual and mean learning scores (word–picture matching accuracy) over the 5 training–testing sessions.

constructed in MarsBaR. The mean beta estimates from the single-subject GLM analysis for each of these ROIs were extracted with MarsBaR and further processed in SPSS. Repetition suppression to tone was estimated as described for the whole-brain analysis and analyzed in a 2 × 2 repeated-measures ANOVA with tone (repeat and change) and session (pre-training and post-training) as factors. Pairwise Pearson's correlations between repetition suppression to tone in the different ROIs and the final learning score (generalization test) were estimated to investigate whether individual variability in learning correlated with the size of the repetition suppression to tone effect.

Psychophysiological Interaction Analyses

To investigate changes in functional connectivity induced by learning, we performed psychophysiological interaction (PPI) analyses in SPM8 for a number of seed regions. We selected the seed regions (volume of interest; VOIs) that, according to the literature, are involved at different stages of pitch processing: the IC, MGB, and HG (Javad et al. 2014). Since we were also interested in top-down connectivity, we also included the left IFG as a VOI. These were anatomically defined as described in the ROI section. We first estimated the physiological factor by extracting the first eigenvariate of the time courses for the voxels within the ROI. The psychological factor was then defined as the repetition suppression to tone effect (tone change conditions > tone repeat conditions) and was used to estimate the interaction term (seed region × effect of tone repetition). Finally, a new GLM analysis was performed for each participant and VOI, with the 16 experimental conditions, the physiological, the psychological, and the PPI terms as regressors, and the 24 realignment parameters as regressors of no interest. The individual contrast images for the interaction terms were then entered in one-sample *t*-tests at the second level for group inferences to test for the functional connectivity difference between the 2 experimental conditions (tone change vs. tone repeat).

Results

Behavioral Results

The behavioral analysis of participants' learning scores (percentage correct) yielded a significant effect of session [$F_{1.605, 49.750}$ = 97.187, $P < 0.001$ (Greenhouse–Geisser-corrected), $\eta_p^2 = 0.758$]. All the post hoc comparisons were highly significant (Table 3), indicating that participants improved over the course of training. Participants also improved in Pitch Discrimination accuracy [$t_{(30)}$ = −4.219, $P < 0.001$] and Pitch Identification accuracy [$t_{(30)} = −4.244$, $P < 0.001$] after training compared with before (Table 4). Although

all participants improved, as expected, their performance varied considerably as indicated by their learning trajectories (Fig. 3).

The correlation between the final Dutchinese learning score (generalization) and participants' pre-training Pitch Discrimination and Identification accuracy was highly significant ($r = 0.603$, $P < 0.001$ and $r = 0.770$, $P < 0.001$, respectively; Table 5). No correlation was found between the final Dutchinese learning score and participants' Backward Digit Span score, Raven's score, music education duration, music education onset, or self-reported motivation (Table 6). We can therefore conclude that learning attainment was specific to sharpening participants' tone processing abilities rather than the result of general cognitive or musical abilities.

Imaging Results

None of the participants understood the tone repetition manipulation in the scanner, as was evident from their responses to the post-scanning questionnaire. Instead, they were all convinced that they were performing a task about sound amplitude changes and had difficulties retrieving the words or the number of speakers they heard while in the scanner.

Whole Brain

Whole-brain comparison results are summarized in Table 7. For the pre-training session, whole-brain comparisons yielded significant repetition suppression effects to tone in the bilateral IFG (Fig. 4). More specifically, the pars opercularis (POp) and pars triangularis (PTr) in the left IFG as well as the POp and precentral gyrus in the right hemisphere were significantly less activated in trials where the tone was repeated compared with trials where the tone had changed. Overall, we did not observe repetition suppression to other acoustic stimulus dimensions (voice and CVC) and no repetition enhancement effects. For the post-training session, we did not find any significant effect for repetition suppression or enhancement to tone. The only whole-brain effect that was significant in the post-training session was a repetition suppression effect to voice in the precuneus.

A comparison across sessions indicated an increase in activation to tone repetitions in the post-training session compared with the pre-training session, but this was not specific to tone; a conjunction analysis showed that the same areas, including the bilateral POp, the left supplementary motor area (SMA), the left thalamus, and the IC, were also more active in the post-training session for repetition of voice and CVC (Table 7 and Fig. 5). Thus, the absence of post-training repetition suppression to tone seemed to be driven by an overall activation increase in response to any repeated acoustic information (tone, voice, or CVC). A similar conjunction analysis was performed on post > pre-training activation to tone, voice, and CVC change. It revealed more activation for post- compared with pre-training in the anterior cingulate cortex (ACC), mid cingulum, and thalamus (Fig. 5).

Given that large between-session changes in the amplitude of fMRI activation can occur due to global effects (Zandbelt et al. 2008; Raemaekers et al. 2012), we ran additional analyses targeting repetition suppression in each session separately. Repetition suppression to tone was estimated by contrasting the TcVreCre condition with the TreVreCre condition for each subject in each session. A one-sample *t*-test was then run on these contrasts for each session with subject as the random factor. No significant activation cluster was found for either session. However, bearing in mind that our training required that participants not only learn the different tone contours but also the different CVC

Table 3 Post hoc comparisons for the effect of session

Final learning score	Comparisons				
	Session 1	Session 2	Session 3	Session 4	Session 5
Mean difference					
Session 1	–	−16.829**	−33.887**	−40.234**	−45.638**
Session 2	16.829**	–	−17.057**	−23.405**	−28.809**
Session 3	33.887**	17.057**	–	−6.348**	−11.751**
Session 4	40.234**	23.405**	6.348**	–	−5.404*
Session 5	45.638**	28.809**	11.751**	5.404*	–

**P < 0.001, *P = 0.005, P Bonferroni-corrected for multiple comparisons.

Table 4 Paired T-tests on tone discrimination and identification accuracy

Accuracy	Pre-training		Post-training		n	95% CI for mean difference	r	t	df
	M	SD	M	SD					
Discrimination	93.99	6.34	96.34	5.02	31	−3.49, −1.21	0.87**	−4.21**	30
Identification	64.06	16.73	72.04	22.77	31	−11.82, −4.14	0.90**	−4.24**	30

**P < 0.001.

Table 5 Correlations between participants' final learning score and tone perception measures

Measure	Pre tone discrimination	Pre tone identification	Post tone discrimination	Post tone identification
Learning score	0.603**	0.770**	0.603**	0.805**
Pre tone discrimination		0.546**	0.876**	0.586**
Pre tone identification			0.599**	0.904**
Post tone discrimination				0.613**

**P < 0.001.

Table 6 Correlations between participants' final learning score and control measures

Measure	Backward DS	Raven	Length music education	Onset music education	Motivation
Learning score	0.189	0.150	0.250	0.218	0.280
Backward DS		0.217	−0.036	−0.035	0.303
Raven			−0.042	−0.181	−0.091
Music education				0.647*	0.179
Onset music education					−0.080

*P < 0.001.

words in which the contours were embedded, we also looked at the condition that would be more sensitive to learning, that is, changes in tone and CVC: TcVreCc vs. TreVreCre. We modeled this specific contrast at the subject level for each session and performed a one-sample t-test at the group level for each session. The results revealed no significant activation in the pre-training session, but significant repetition suppression to tone and CVC in the post-training session, in the left precentral gyrus, PTr, and POp (Table 8 and Fig. 6). These more specific comparisons indicate that, after training, participants became more efficient in processing changes in tone and CVC information, both of which having been crucial for Dutchinese word learning. This was not driven by CVC change alone, since the contrast between TreVreCc and TreVreCre did not yield any significant results in either session.

Repetition Suppression Effect Along the Auditory Pathways
The ROI analysis aimed to increase sensitivity in detecting repetition suppression effects in brain areas that have been reported to process acoustic changes. The repeated-measures ANOVA on the extracted beta estimates revealed a significant effect of session with overall more activation to the stimuli on the post-training compared with the pre-training scanning session in the left and right IFG (POp, PTr, and Pars Orbitalis), the right STG, and the thalami. A significant effect of Condition (tone change vs. tone repeat) was found in the bilateral IFG and thalami, with more activation for tone change compared with the tone repeat condition (Fig. 7A,B). A significant Session × Condition interaction was found in the right HG and right POp. The interaction was driven by a large repetition suppression effect in the pre-training session and a much weaker effect in the post-training session.

There was a significant correlation between participants' final learning score and repetition suppression to tone in the left IFG (r = 0.432, P = 0.014 for POp and r = 0.424, P = 0.016 for PTr) after training (Fig. 7D). Interestingly, participant's repetition suppression to tone in the left POp correlated with their final learning score even in the pre-training session (r = 0.361, P = 0.042; Fig. 7C). This correlation seems to be driven by the fact that good learners' left POp deactivated more when a tone was repeated, compared with less good learners in the pre-training session (correlation between the learning score and activation to tone repetition: r = −0.384, P = 0.03). Apart from a marginal positive correlation between the learning score and repetition suppression in the right HG (r = 0.324, P = 0.071) after training, no other correlations reached significance.

Functional Connectivity Along the Auditory Pathway
The purpose of the PPI analyses was to explore connectivity changes among auditory language areas as a result of tone learning. We therefore focused on areas that are part of the pitch processing pathway ranging from subcortical (IC) to higher-order cortical brain regions (IFG). With the contrast of tone change vs. tone repeat as the psychological factor, no cluster survived the whole-brain comparison in the pre-training session. However, in the post-training session, we found a significant increase in connectivity between the right HG and left POp with tone repetition (Fig. 8, peak local maximum [−36, 18, 20], P = 0.021, FWE-corrected). That is, after training had taken place, the strength of the association between activity in the right HG and left POp was greater on tone repetition trials than on tone change trials. This post-training connectivity between the right HG and the left POp, however, did not correlate with participants' learning

Table 7 Whole-brain analysis results

Contrast	Region	H	No. of voxels	MNI coordinates			T	Z	Cluster p(FWE)
				x	y	z			
Pre-training session									
Tone change > tone repeat	Precentral gyrus	R	1447	44	8	20	4.83	4.77	<0.001
	Pars opercularis			40	6	32	4.63	4.57	
				48	14	30	4.22	4.18	
	Pars opercularis	L	537	−48	10	22	4.65	4.6	0.001
	Pars triangularis			−36	28	14	4.57	4.52	
	Pars opercularis			−40	6	28	3.97	3.94	
Voice change > tone repeat	n.s.								
CVC change > tone repeat	n.s.								
Post-training session									
Tone change > tone repeat	n.s.								
Voice change > tone repeat	Precuneus	L	430	−8	−60	26	4.40	4.36	0.002
				−10	−54	16	4.18	4.14	
				−4	−60	16	3.96	3.92	
CVC change > tone repeat	n.s.								
Post-training > pre-training (conjunction)									
Repetition	Pars opercularis	L	1016	−48	12	14	4.85	4.79	<0.001
	Precentral gyrus			−46	8	30	4.76	4.7	
				−40	4	36	4.63	4.58	
	Supplementary motor area	L	709	−2	24	38	4.5	4.45	<0.001
	Mid cingulum			−4	16	42	4.48	4.44	
				−10	12	46	4.4	4.36	
	Pars triangularis	R	334	44	22	26	4.14	4.1	0.007
	Pars opercularis			46	30	26	3.92	3.88	
				48	14	32	3.59	3.57	
	~Inferior colliculus	L	265	−4	−30	0	3.96	3.92	0.016
	~Thalamus			−12	−42	4	3.72	3.69	
				2	−26	10	3.3	3.28	
Change	Mid cingulum	R	422	0	22	40	4.25	4.21	0.002
	Anterior cingulum	L		−6	26	36	4.19	4.15	
		L		−4	30	26	3.94	3.9	
	Thalamus	R	233	2	−10	8	4.1	4.06	0.025
		L		0	−18	10	3.78	3.75	
				−14	−14	12	3.27	3.25	

Note: Region labels were provided by the AAL Atlas using the MNI coordinates.

~, approximate location; n.s., not significant.

attainment. No other area showed significant connectivity changes in the post-training session.

Discussion

We investigated individual variation in non-native tone learning performance by measuring fMRI adaptation to tones before and after administering a multisession tone training procedure. Our behavioral results demonstrate that Dutch native listeners were able to learn to associate words that differed minimally in pitch contour with meaning, since their performance improved significantly with training. Based on participants' post-training improvement in tone discrimination and identification tasks, we can be confident that these results do not reflect simple associative learning, but are specific to learning the non-native contrast. At the same time, we observed large individual variability in the participants' learning trajectories, replicating previous studies that used a similar paradigm (Wong et al. 2007; Chandrasekaran et al. 2012). The fact that the participants' final learning scores correlated positively with their ability to accurately discriminate and identify tone patterns before

training supports the notion of pre-existing differences in learning aptitude, such that the learners who processed tone contours more efficiently benefited more from the tone training.

Overall, our Dutch native listeners showed repetition suppression to non-native tones in the bilateral IFG, including the right precentral gyrus and bilateral POp and PTr, prior to training. This was in accordance with our expectations, since bilateral IFG deactivation has been consistently reported in studies of fMRI adaptation to repeated auditory information. With respect to spoken language, IFG deactivation has been found in spoken sentence repetition (Hasson et al. 2006), phonological feature repetition (Vaden et al. 2010), in repetition of non-native consonants (Myers and Swan 2012), and with repetition of phonemes of the same phonetic category (Myers et al. 2009). A linear decrease (repetition suppression) in these areas is also observed when musical notes are repeated in short melodies (Brown et al. 2013) or when the perceived voice gender is repeated (Charest et al. 2013). It thus seems that IFG activity is associated with perception of acoustic information, especially in cases where explicit judgments on this information are required (Hasson et al. 2006).
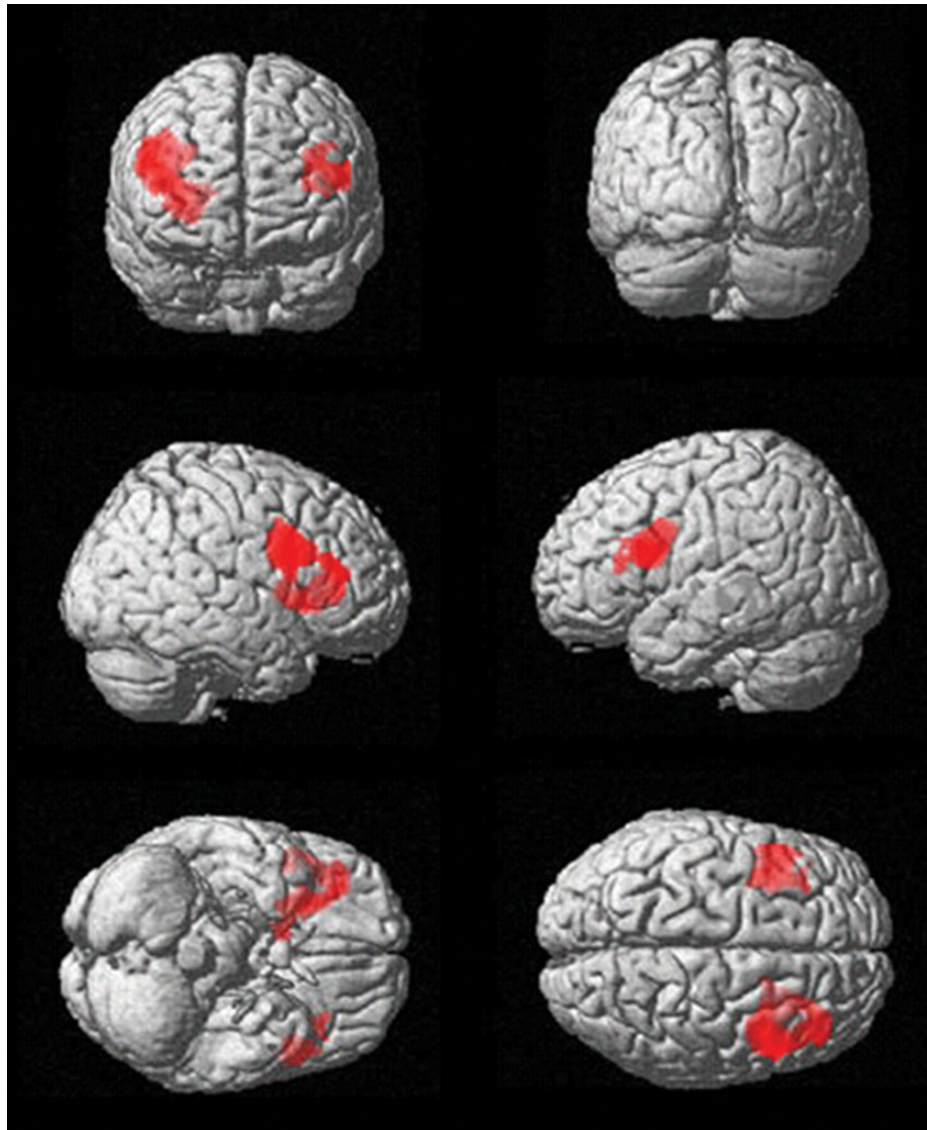
Figure 4. Repetition suppression to tone in the pre-training session. Significantly less activation with tone repetition was found in the left POp and PTr, and in the right POp and precentral gyrus (uncorrected $P < 0.001$, FWE cluster-corrected).

It is possible that our participants used their knowledge of intonation and prosody while processing the non-native pitch contours. Although Standard Dutch does not use pitch at the lexical level, it does use rising and falling pitch contours at the suprasegmental prosodic level ('t Hart 1998). A recent meta-analysis has shown that the bilateral PTr is activated when processing affective prosody and the bilateral POp for linguistic prosody, while the right precentral gyrus is involved in both (Belyk and Brown 2014). It could therefore be the case that, upon listening to these tones for the first time, Dutch listeners interpreted them as prosodic contours, yielding larger repetition suppression in the right IFG. This would be in accordance with lateralization patterns in prosodic processing (Rota et al. 2009; Witteman et al. 2012; Belyk and Brown 2014).

Importantly, repetition suppression in the left IFG, and particularly the POp, correlated positively with tone learning performance, such that individuals who were better learners of tones showed larger repetition suppression to tone in this area even before training. Our findings thus support the hypothesis

that variation in sound-learning aptitude stems in part from the fact that individuals differ in how efficiently they encode and process non-native sound contrasts. Although all learners improved significantly with training, converging fMRI (pre-training repetition suppression to tone) and behavioral data (pre-training tone identification accuracy) demonstrate that they did not start off at the same level.

Consistent with our repetition suppression (i.e., deactivation) findings, activation in response to non-native sounds in the left IFG has been shown to correlate negatively with sound-learning performance (Golestani and Zatorre 2004; Myers and Swan 2012). Previous findings have been interpreted in a speculative manner, with accounts alluding to verbal working memory or subvocal rehearsal as the potential underlying mechanisms of left IFG activation patterns. Assuming that they lack clear representations of lexical sounds, less successful learners would rely on encoding any acoustic information available and keeping it online. This would take up more verbal memory resources to support their performance compared with successful learners. Although
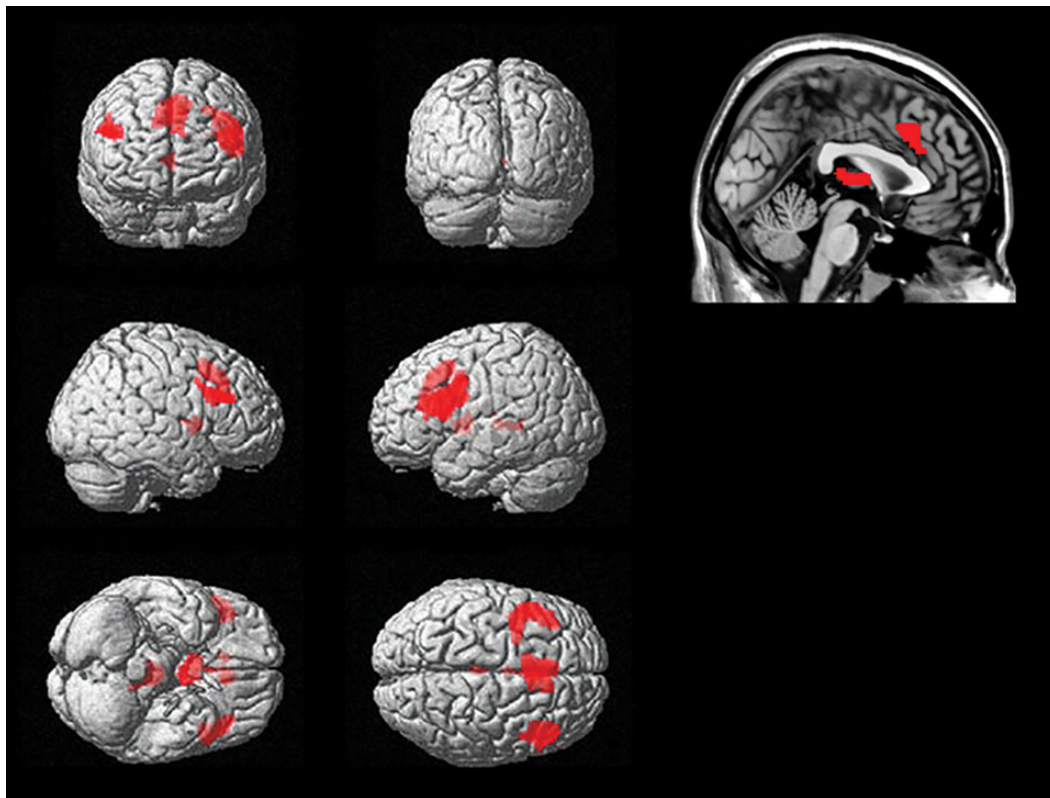
Figure 5. Conjunction analysis results (uncorrected P < 0.001, FWE cluster-corrected). Left: conjunction analysis of post-training versus pre-training tone repetition, voice repetition, and CVC repetition. The bilateral POp, the left SMA, left thalamus, and IC were more active for any acoustic repetition in the post-training session. Right: sagittal view of conjunction analysis for post-training versus pre-training tone change, voice change, and CVC change. Increased activation in the ACC, mid cingulum, and thalamus to changing acoustic information after training.

Table 8 Significant clusters revealed by the one-sample *t*-test on repetition suppression to tone and CVC in the post-training session

Contrast	Region	H	No. of voxels	MNI coordinates			T	Z	Cluster *p*(FWE)
				x	*y*	*z*			
Post-training session									
TcVreCc > TreVreCre	Precentral gyrus	L	342	−44	2	34	4.83	4.14	0.003
	Pars triangularis			−42	32	20	4.75	4.08	
	Pars opercularis			−50	10	28	4.41	3.85	

Note: Region labels were provided by the AAL Atlas using the MNI coordinates.

we cannot completely exclude such an interpretation, it seems unlikely in our case because we did not observe a correlation between our behavioral verbal working memory measure and learning performance. Instead, a more favorable interpretation is that the left POp is involved in controlling and deciding on relevant abstract stimulus representations (Hasson et al. 2007; Myers et al. 2009; Myers and Swan 2012), thereby guiding learning in sensory encoding areas by means of top-down feedback connections. Less successful learners would accordingly need more top-down feedback than successful learners, since they have not yet built efficient representations of the stimuli to inform perception [see also Golestani and Zatorre (2004)].

Contrary to our expectations, repetition suppression to tone did not increase with training, as was evident in the whole-brain full-factorial analysis of pre- and post-training data. Our session-specific analyses, however, did reveal that training induced repetition suppression to repeating combinations of

tones and CVCs in the left ventrolateral prefrontal cortex in the post-training session, regardless of the level of learning attainment. This finding makes sense in the context of our training paradigm, where learning the Dutchinese words required participants to pay attention to both tonal and segmental (CVC) information. The segments of the novel words were acoustically salient and, unlike voice, were highly relevant for learning the words, just like the pitch contours were.

Providing increased sensitivity, the ROI analysis allowed us to detect repetition suppression to tone after training completion. This effect was there for thalamic and bilateral frontal areas. It was, however, smaller compared with the pre-training sessions, mainly due to the increase in the BOLD response for tone repetitions rather than the decrease for tone changes. Overall, activation was higher in the post-training session along the bilateral IFG, the right STG, and thalamus, maybe because the participants had learned to associate the Dutchinese words with meanings
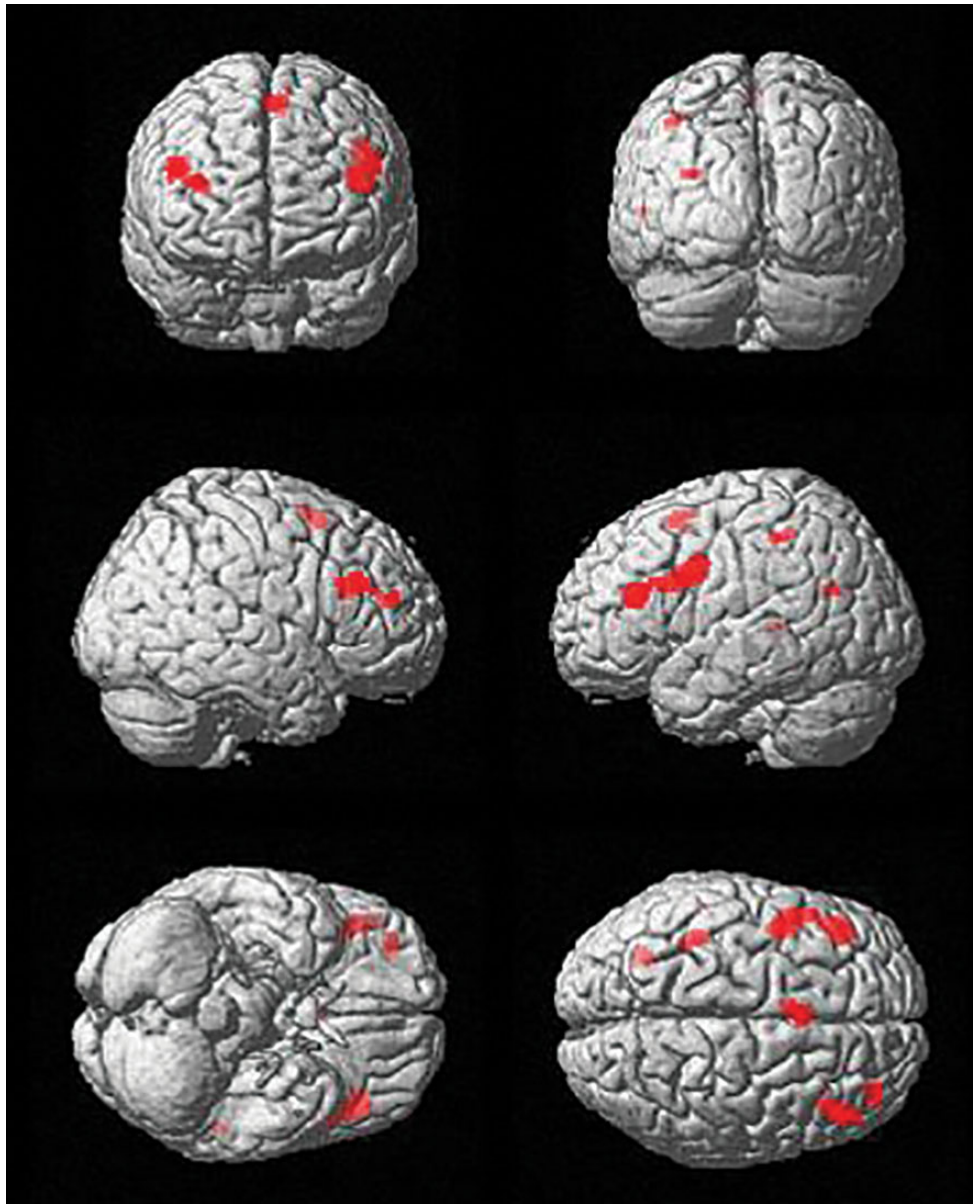
Figure 6. One-sample *t*-test on repetition suppression to tone and CVC in the post-training session. A cluster in the left precentral gyrus, POp, and PTr showed significantly less activation when tone and CVC information was repeated after training (uncorrected *P* < 0.001, FWE cluster-corrected).

over the course of the training. We cannot exclude the possibility that the newly acquired semantic representations of the words might have influenced the brain activity pattern in the post-training session. This could account for the overall higher activation in the post-training session in the bilateral IFG, the right STG, and thalamus. It is unlikely, however, that our results could be explained by changes in the awareness of the stimuli, since post-scanning reports indicate that participants were completely unaware of the tone repetition manipulation, and their recall of the presented words and the number of speakers required a lot of effort and was not always successful.

Our functional connectivity analysis revealed an increase in the strength of association between activation in the right HG and the left POp with tone repetition after training, regardless of learning performance. Although it is difficult to make directionality claims, we speculate that this reflects an increase in

feed-forward connectivity from a basic pitch encoding area, such as the right primary auditory cortex, to higher-order pitch contour representations in the left frontal cortex. As mentioned earlier, the behavioral results suggest that learning has taken place, as evident from the improvement in discrimination and identification of tone patterns across participants. Thus, in the post-training session, all participants must have improved to some extent in encoding pitch information, which preferentially engages the right HG (Luo et al. 2006; Xu et al. 2006; Warrier et al. 2009; McGettigan and Scott 2012). A similar right temporal–left frontal network has been postulated to underlie domain-general pitch processing by Nan and Friederici (2013). They suggest that the right auditory cortex does the initial pitch acoustic processing while the left IFG does the more cognitive and decision-related processing (Nan and Friederici 2013). The fact that we observe what appears to be feed-forward instead
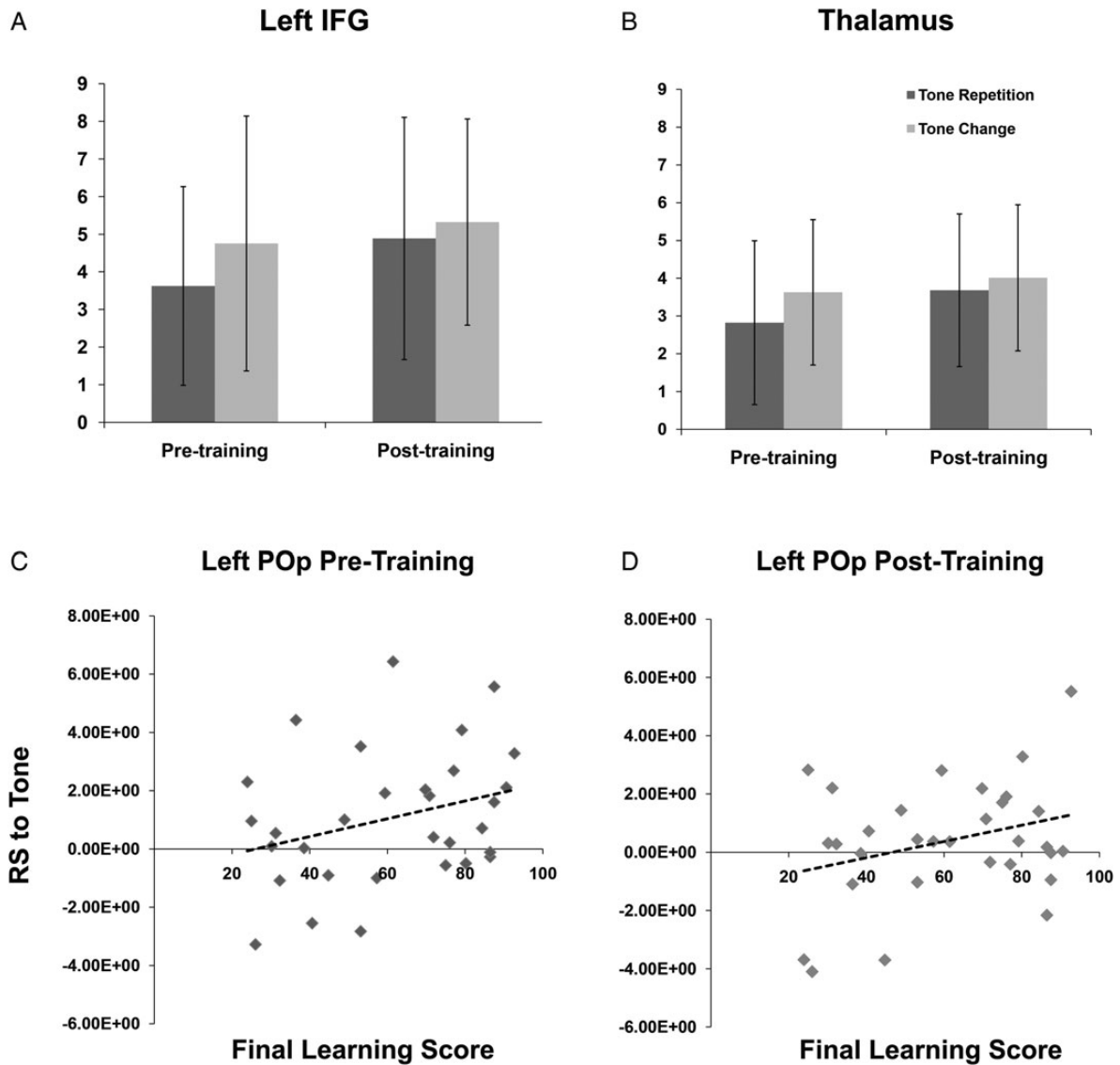
Figure 7. Mean activations (arbitrary scale) to tone repetition and tone change (A) in the left IFG and (B) in the thalamus, pre- and post-training. Error bars denote 1 SE around the mean. (C) Scatter plot of repetition suppression (RS) to tone and final learning score in the left POp for the pre-training and (D) post-training session.
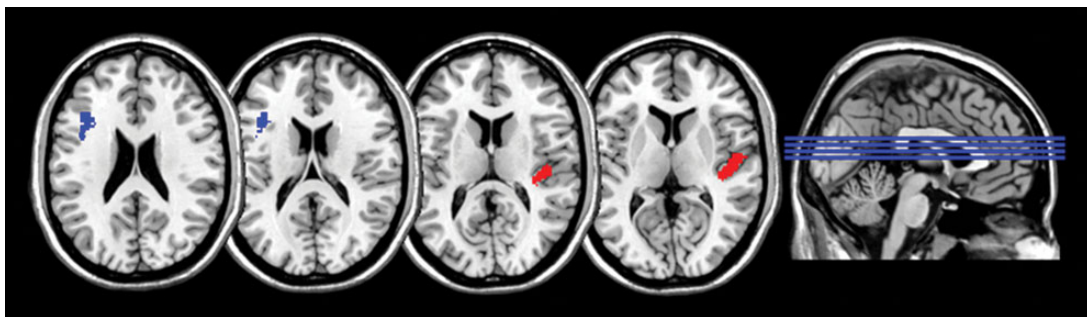

Figure 8. Multislice view of the cluster in the left POp (blue) that showed a significant increase in connectivity with the right HG (red) seed region in the PPI analysis (uncorrected $P < 0.001$, FWE cluster-corrected).

of feedback connectivity can be attributed to the task participants performing in the scanner (i.e., the amplitude change detection task). This required forwarding accurate acoustic information from sensory areas to higher-order representation and decision areas. In this context, feedback connectivity is rendered unnecessary, which probably explains why the strength of connectivity between these areas did not correlate with learning performance.

The absence of adaptation effects in the temporal lobes, otherwise often reported in auditory fMRI adaptation (Hasson et al. 2006; Rauschecker et al. 2008; Hu et al. 2013), might be due to our design. We used a slow event-related design with a long lag between repetitions (~14 s), which may have been too long for more sensory-related repetition suppression effects to arise (Grill-Spector et al. 2006). It is also possible that there was repetition suppression to tone in the primary and secondary auditory cortices, but it might have been sensitive to the number of repetitions. With only 4 tones available, we could not avoid repeating them multiple times across the experiment. As a consequence of this, activation levels in sensory areas might have reached saturation. Finally, there is the possibility that these areas showed repetition suppression, but that it was not large enough to survive whole-brain comparisons. Our ROI analysis, however, argues against this. Myers and Swan (2012) also did not find changes in STG after categorical phonetic training and attribute this absence to the fact that training was very short. Changes in temporal areas dedicated to more sensory processing may require long-term exposure to new sounds. Given that such changes should occur through top-down feedback from frontal areas, the patterns of IFG activation we report here could be an indication of establishing the first stage of the sound-learning process.

We knew from Chandrasekaran et al. (2012) that even basic pitch encoding structures, such as the IC, contribute to non-native sound learning. Now, we also have evidence that higher-order cortical structures, such as the left IFG, are important for learning performance. It is our hope that future studies with longitudinal training paradigms can investigate long- term sound learning and shed more light onto the role of fronto-temporal as well as subcortical sound encoding areas in this process.

To conclude, we trained Dutch native speakers in non-native Mandarin tones over 5 separate sessions. fMRI adaptation data to tones were acquired before and after training to assess tone processing efficiency and how it changes with learning. Participants showed repetition suppression to tones in the bilateral IFG before training. Training induced repetition suppression to the combination of tones and CVC segments, the two relevant sources of acoustic information for learning. There was no whole-brain repetition suppression effect to tone post-training, but an increased general sensitivity to any repeated acoustic information. This increased sensitivity could be due to increased feed-forward connectivity between right auditory and left frontal regions. While all participants showed behavioral improvement, they started and ended the training at different levels, with substantial individual variation in their learning scores. Some individuals were thus better than others in learning non-native tones. We attribute their improved learning performance to more efficient processing of tones, as revealed by the correlation between repetition suppression in the left IFG and learning performance. Strikingly, this correlation was there even before training began. This suggests that individual differences in speech-learning aptitude reflect, at least in part, differences in neuronal processing efficiency, in particular in the left IFG.

Funding

Notes

References

Ahissar M, Nahum M, Nelken I, Hochstein S. 2009. Reverse hierarchies and sensory learning. Philos Trans R Soc Lond B Biol Sci. 364:285–299.

Bajo VM, Nodal FR, Moore DR, King AJ. 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. Nat Neurosci. 13:253–260.

Belyk M, Brown S. 2014. Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. Soc Cogn Affect Neurosci. 9:1395–1403.

Best CT, McRoberts GW, Goodell E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. J Acoust Soc Am. 109:775.

Brown RM, Chen JL, Hollinger A, Penhune VB, Palmer C, Zatorre RJ. 2013. Repetition suppression in auditory-motor regions to pitch and temporal structure in music. J Cogn Neurosci. 25:313–328.

Chandrasekaran B, Kraus N, Wong PCM. 2012. Human inferior colliculus activity relates to individual differences in spoken language learning. J Neurophysiol. 107:1325–1336.

Chandrasekaran B, Sampath PD, Wong PCM. 2010. Individual variability in cue-weighting and lexical tone learning. J Acoust Soc Am. 128:456–465.

Charest I, Pernet C, Latinus M, Crabbe F, Belin P. 2013. Cerebral processing of voice gender studied using a continuous carry-over fMRI design. Cereb Cortex. 23:958–966.

Flege JE, Bohn O-S, Jang S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. J Phon. 25:437–470.

Flege JE, Yeni-Komshian GH, Liu S. 1999. Age constraints on second-language acquisition. J Mem Lang. 41:78–104.

Golestani N, Zatorre RJ. 2009. Individual differences in the acquisition of second language phonology. Brain Lang. 109:55–67.

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

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci. 10:14–23.

Gubian M. 2011. Functional Data Analysis for Phonetics Research. 1–4.

Hanulíková A, Dediu D, Fang Z, Bašnaková J, Huettig F. 2012. Individual differences in the acquisition of a complex L2 phonology: a training study. Lang Learn. 62:79–109.

't Hart J. 1998. Intonation in Dutch. In: Hirst D, Di Cristo A, editors. Intonation Systems: A Survey of Twenty Languages. New York: Cambridge University Press. p. 96–111.

Hasson U, Nusbaum HC, Small SL. 2006. Repetition suppression for spoken sentences and the effect of task demands. J Cogn Neurosci. 18:2013–2029.

Hasson U, Skipper JI, Nusbaum HC, Small SL. 2007. Abstract coding of audiovisual speech: beyond sensory representation. Neuron. 56:1116–1126.

Hayasaka S, Nichols TE. 2003. Validating cluster size inference: random field and permutation methods. Neuroimage. 20:2343–2356.

Hu X, Ackermann H, Martin Ja, Erb M, Winkler S, Reiterer SM. 2013. Language aptitude for pronunciation in advanced

second language (L2) learners: behavioural predictors and neural substrates. Brain Lang. 127:366–376.

Javad F, Warren JD, Micallef C, Thornton JS, Golay X, Yousry T, Mancini L. 2014. Auditory tracts identified with combined fMRI and diffusion tractography. Neuroimage. 84:562–574.

Luo H, Ni J-T, Li Z-H, Li X-O, Zhang D-R, Zeng F-G, Chen L. 2006. Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. Proc Natl Acad Sci USA. 103:19558–19563.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage. 19:1233–1239.

McGettigan C, Scott SK. 2012. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? Trends Cogn Sci. 16:269–276.

Mühlau M, Rauschecker JP, Oestreicher E, Gaser C, Röttinger M, Wohlschlägera M, Simon F, Etgen T, Conrad B, Sander D. 2006. Structural brain changes in tinnitus. Cereb Cortex. 16:1283–1288.

Myers EB, Blumstein SE, Walsh E, Eliassen J. 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci. 20:895–903.

Myers EB, Swan K. 2012. Effects of category learning on neural sensitivity to non-native phonetic categories. J Cogn Neurosci. 24:1695–1708.

Nan Y, Friederici AD. 2013. Differential roles of right temporal cortex and Broca's area in pitch processing: evidence from music and Mandarin. Hum Brain Mapp. 34:2045–2054.

Poser Ba, Versluis MJ, Hoogduin JM, Norris DG. 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. Magn Reson Med. 55:1227–1235.

Race EA, Shanker S, Wagner AD. 2009. Neural priming in human frontal cortex: multiple forms of learning reduce demands on the prefrontal executive system. J Cogn Neurosci. 21:1766–1781.

Raemaekers M, du Plessis S, Ramsey NF, Weusten JMH, Vink M. 2012. Test-retest variability underlying fMRI measurements. Neuroimage. 60:717–727.

Rauschecker AM, Pringle A, Watkins KE. 2008. Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. Hum Brain Mapp. 29:1231–1242.

Rossion B, Pourtois G. 2004. Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. Perception. 33:217–236.

Rota G, Sitaram R, Veit R, Erb M, Weiskopf N, Dogil G, Birbaumer N. 2009. Self-regulation of regional cortical activity using real-time fMRI: the right inferior frontal gyrus and linguistic processing. Hum Brain Mapp. 30:1605–1614.

Schönwiesner M, Novitski N, Pakarinen S, Carlson S, Tervaniemi M, Näätänen R. 2007. Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. J Neurophysiol. 97:2075–2082.

Segaert K, Weber K, de Lange FP, Petersson KM, Hagoort P. 2013. The suppression of repetition enhancement: a review of fMRI studies. Neuropsychologia. 51:59–66.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15:273–289.

Vaden KI, Muftuler LT, Hickok G. 2010. Phonological repetition-suppression in bilateral superior temporal sulci. Neuroimage. 49:1018–1023.

Ventura-Campos N, Sanjuán A, González J, Palomar-García M-Á, Rodríguez-Pujadas A, Sebastián-Gallés N, Deco G, Ávila C. 2013. Spontaneous brain activity predicts learning ability of foreign sounds. J Neurosci. 33:9295–9305.

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.

Warrier C, Wong P, Penhune V, Zatorre R, Parrish T, Abrams D, Kraus N. 2009. Relating structure to function: Heschl's gyrus and acoustic processing. J Neurosci. 29:61–69.

Witteman J, Van Heuven VJP, Schiller NO. 2012. Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. Neuropsychologia. 50:2752–2763.

Wong PCM, Perrachione TK. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. Appl Psycholinguist. 28:565–585.

Wong PCM, Perrachione TK, Parrish TB. 2007. Neural characteristics of successful and less successful speech and word learning in adults. Hum Brain Mapp. 28:995–1006.

Xu Y, Gandour J, Talavage T, Wong D, Dzemidzic M, Tong Y, Li X, Lowe M. 2006. Activation of the left planum temporale in pitch processing is shaped by language experience. Hum Brain Mapp. 27:173–183.

Yan J, Zhang Y, Ehret G. 2005. Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. J Neurophysiol. 93:71–83.

Zandbelt B, Gladwin TE, Raemaekers M, van Buuren M, Neggers SF, Kahn RS, Ramsey NF, Vink M. 2008. Within-subject variation in BOLD-fMRI signal changes across repeated measurements: quantification and implications for sample size. Neuroimage. 42:196–206.

Zatorre RJ. 2013. Predispositions and plasticity in music and speech learning: neural correlates and implications. Science. 342:585–589.