NeuroImage 192 (2019) 76–87

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# Speech processing and plasticity in the right hemisphere predict variation in adult foreign language learning

Zhenghan Qi [a,b,*], Michelle Han [b], Yunxin Wang [b,c], Carlo de los Angeles [b], Qi Liu [b,c], Keri Garel [b], Ee San Chen [b], Susan Whitfield-Gabrieli [d], John D.E. Gabrieli [b,e], Tyler K. Perrachione [f]

[a] Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, 19711, USA
[b] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[c] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China
[d] Department of Psychology, Northeastern University, USA
[e] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[f] Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA, 02215, USA

ARTICLE INFO

*Keywords:*
fMRI
Individual differences
Language learning
Prediction
Resting state connectivity
Speech perception

ABSTRACT

Foreign language learning in adulthood often takes place in classrooms where learning outcomes vary widely among students, for both initial learning and long-term retention. Despite the fundamental role of speech perception in first language acquisition, its role in foreign language learning outcomes remains unknown. Using a speech discrimination functional magnetic resonance imaging (fMRI) task and resting-state fMRI before and after an intensive, classroom-based, Mandarin Chinese course, we examined how variations in pre-training organization and pre-to-post reorganization of brain functions predicted successful language learning in male and female native English-speakers. Greater pre-training activation in right inferior frontal gyrus (IFG) to Mandarin speech was associated with better Mandarin attainment at the end of the course. After four weeks of class, learners showed overall increased activation in left IFG and left superior parietal lobule (SPL) to Mandarin speech, but in neither region was variation related to learning outcomes. Immediate attainment was associated with greater pre-to-post reduction of right IFG activation to Mandarin speech but also greater enhancement of resting-state connectivity between this region and both left IFG and left SPL. Long-term retention of Mandarin skills measured three months later was more accurately predicted by models using features of neural preparedness (pre-training activation) and neural plasticity (pre-to-post activation change) than models using behavior preparedness and plasticity features (pre-training speech discrimination accuracy and Mandarin attainment, respectively). These findings suggest that successful holistic foreign language acquisition in human adulthood requires right IFG engagement during initial learning but right IFG disengagement for long-term retention of language skills.

## 1. Introduction

To acquire a language, both infants at home and adults in the classroom start by learning speech sound categories and conclude with sophisticated language skills that seamlessly integrate phonology, semantics, syntax, and orthography. Infants' early perception of native speech sounds has been associated with longitudinal growth in their native language (Rivera-Gaxiola et al., 2005; Kuhl et al., 2008). Although the capacity to discriminate non-native speech sounds fades during early development, some degree of non-native speech perception is preserved in adults and varies substantially across individuals (Sebastián-Gallés and Díaz, 2012). In adulthood, better behavioral performance in discriminating non-native speech sounds is related to greater success learning foreign language vocabulary and phonology (Wong and Perrachione, 2007; Chandrasekaran et al., 2010; Silbert et al., 2015), suggesting that non-native speech perception abilities may act as a "gatekeeper" to the initial stages of foreign language learning.

Prior laboratory-based training studies have focused on the neural predictors and plasticity of only circumscribed elements of language learning (e.g., vocabulary or syntax) or laboratory-based training with a miniature artificial language. Even prior to training, more robust hemodynamic responses to novel phonetic features distinguished more successful from less successful learners of an artificial tonal language vocabulary (Wong et al., 2007; Yang et al., 2014). Furthermore, facility

* Corresponding author. University of Delaware, 125 E. Main St., Rm. 103, Newark, DE, 19711, USA.
*E-mail address:* zqi@udel.edu (Z. Qi).

https://doi.org/10.1016/j.neuroimage.2019.03.008
Received 15 October 2018; Received in revised form 20 February 2019; Accepted 4 March 2019
Available online 8 March 2019

with speech-sound processing may influence broader aspects of language learning, including syntax: When learning the grammar of a miniature artificial language, the similarity of its sound structure to learners' native-language phonology determined the native-likeness of fMRI responses to novel syntactic structures (Finn and Hudson Kam, 2008; Finn et al., 2013). Likewise, stronger event-related potentials (ERP) responses to speech phonetics were observed in individuals who were able to successfully extract embedded syntactic rules in an artificial language compared to those who failed to do so (Mueller et al., 2012). These laboratory-based studies all suggest that speech perception abilities are foundational for learning not only vocabulary but also higher-level aspects of a new language such as syntax. However, it remains unknown how the interplay between speech perception and language learning – and particularly the neural mechanisms underlying these processes – determines the long-term retention of holistic, real-world foreign language skills.

Classroom-based training and real-world learning expose learners to new and complex rules in all aspects of language. Moreover, adults' learning success in the real world is determined by their ability to deploy holistic language skills in high-level comprehension and production tasks. Research has shown that adults differ markedly in immediate attainment and long-term retention of foreign language skills (Robinson, 2001; Murtagh and van der Slik, 2004; Kormos and Sáfár, 2008), raising the question of how individual differences in neural preparedness and neural plasticity lead to different long-term learning outcomes.

No study has yet examined the neural underpinnings of the relations between speech perception and successful naturalistic foreign language learning. The few studies of real-world language learning have focused primarily on anatomical changes following learning, finding increased white-matter connectivity between left and right hemispheres and growth of cortical thickness and grey matter volume in left-hemisphere language-related areas after months of language training (Mårtensson et al., 2012; Schlegel et al., 2012; Stein et al., 2012). With respect to brain function, a positive association has been found between functional connectivity measured before training and second-language reading and lexical-retrieval skills measured after a 12-week, classroom-based French course (Chai et al., 2016). Yet, the functional consequence of these structural and resting-state differences has not been explored.

Research on speech perception has revealed a functional dissociation in the hemispheric processing of speech sounds. The left hemisphere is primarily sensitive to phonological differences and is actively involved in speech perception by native speakers, whereas the right hemisphere is primarily sensitive to acoustic differences, with naïve listeners of a language tending to show right-hemisphere lateralization when processing novel speech sounds (Hsieh et al., 2001; Gandour et al., 2002, 2004; Wong et al., 2004; Luo et al., 2006; Xu et al., 2006; Zatorre and Gandour, 2008; Myers et al., 2009; Friederici, 2011). Here, we investigated how pre-training hemispheric organization and post-training hemispheric reorganization for speech processing are functionally related to successful classroom-based language learning in the real world.

In the present study, we examined both resting-state and task-based fMRI before and after classroom-based Mandarin training. We focused on the plasticity of neural sensitivity to Mandarin lexical tones, a set of novel phonological features completely absent from English. We investigated how the change of hemispheric involvement in lexical tone processing and the change of cross-hemispheric connectivity during rest were related to foreign-language learning outcomes. The capacity for retaining proficiency after training is a critical challenge for real-world language learners. No previous study has explored neural predictors of language retention months after the training. We determined how successful long-term retention of Mandarin Chinese as a second language could be predicted based on *neural preparedness* (pre-training neural sensitivity to speech sounds) and *neural plasticity* (changes in neural sensitivity over the course of learning).

## 2. Materials and methods

### 2.1. Participants

Twenty-four native speakers of American English (8 females, 16 males; mean age = 23.2 years, SD = 3.68; 20 right-handed, 4 left-handed) were recruited via public job post websites. Because the four left-handed learners' behavioral and neural data were within the same range of right-handed learners, we elected to include all learners in the subsequent analysis. Separate analyses with only right-handed participants found the same significant patterns. The project was advertised as a real language course in the greater Boston area. The enrollment process was similar to that of any elective language course at the college level. All volunteers gave informed written consent prior to their participation that was approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology. Participants were screened for MRI compatibility and negative history of neurological diseases. The participants had an average full-scale IQ of 118 (SD = 12.34; range: 92–143) measured by the Kaufman Brief Intelligence Test (KBIT-2, Kaufman and Kaufman, 2004), a high average likely related to self-selection bias in the urban and collegiate location of the study. Participants' foreign language learning aptitude was measured by Modern Language Aptitude Test (MLAT; Carroll and Sapon, 2002), with total score percentile on this test ranging between 10% and 99% (M = 65.3%, SD = 30.3%).

As indicated by the Language History Questionnaire (Linck et al., 2013), all 24 participants were native speakers of American English, of whom 19 were monolingual with no extensive exposure to languages other than English as a child. Five participants reported growing up in a multilingual environment (Polish, Spanish, Hindi, Hebrew, German, or French). Among these, four participants spoke a non-English language fluently and one spoke English exclusively during childhood. The five multilingual participants did not differ from the monolingual participants in age, IQ, pre-training task performance in the scanner (see "Task fMRI experiment" below), or language learning outcomes (see "Language training procedure" below; Supplementary Table 1). Participants also reported studying from 1 to 5 foreign languages in their high school and college classwork. Their overall language skills in their most proficient language, averaged across reading, writing, listening, and speaking, ranged from 1 to 5 (1 being "limited" and 5 being "excellent"; M = 3.15, SD = 1.14). Neither number of foreign languages studied nor self-reported foreign language proficiency was associated with participants' Mandarin learning outcomes ($p$'s > 0.23). Crucially, none of the participants had any prior exposure to either Chinese, any other tonal languages (e.g., Thai), or any other language with a logographic writing system (e.g., Japanese Kanji). Of the 24 learners, 21 completed a questionnaire detailing their music training experiences. Nineteen learners reported taking music lessons, ranging from 1 to 24 years (Mean = 7.9 years; Median = 5 years; SD = 6.9 years). None of these learners reported having perfect pitch. The duration of music training experiences, however, was not associated with language learning outcomes ($p$'s > 0.46).

### 2.2. Language training procedure

Participants underwent an intensive, month-long, half-day, classroom-based Mandarin Chinese learning experience, equivalent to one semester of college-level study. Participants signed up for one of two course sections taught by the same instructor. The course used *The New Practical Chinese Reader Textbook 1* (新实用汉语课本; Liu, 2010) as the main textbook. The course, which took place on the MIT campus during the summer, provided students with intensive training in introductory Modern Standard Mandarin, for 3.5 h per day, 5 days per week over 4 weeks. Daily homework was assigned from the companion workbook and was supplemented with additional material designed by the instructor, which included practice with listening, reading and writing. The instructor collected, graded, and provided feedback on the homework every day to ensure its completion and utility in learning. Throughout the course, participants spent, on average, 62.3 h (range: 52.5–66.5) in the classroom and completed 11 assignments, 10 quizzes, one midterm exam, and one final exam. An anonymous midterm survey (to which 14 of the 24 students responded) showed participants also spent, on average, 2.8 h (range: 1.5–5 h) every day outside of class studying Mandarin.

On the last day of class, students' Mandarin learning attainment was measured by the average score across two proficiency exams: the course's regular final exam developed by the instructor and a standardized proficiency test published by the Chinese government. First, students completed the course's regular final exam, which was composed of speech production, listening, and reading sections. This test was written mostly in Chinese characters, which tapped into students' ability to effectively integrate orthographic information with phonological, semantic, and syntactic information. The final exam was given by the instructor in three parts (written, oral, and listening) each on a scale of 0–100. The grade was calculated by averaging the scores across all three parts. Second, students completed the Chinese Proficiency Test (HSK; 汉语水平考试; www.chinesetest.cn). The HSK is an official, standardized exam used by the Chinese National Office for Teaching Chinese as a Foreign Language to test learners' mastery of Mandarin as a foreign language. The Level 1 Sample Test was used because the tested vocabulary and communication skills best matched the scope of the course (i.e., 60 h of Chinese instruction, mastery of at least 150 Chinese words, and certain grammatical rules). The HSK takes 40 min, contains 40 multiple-choice questions, and is divided evenly between sections that test listening and reading comprehension. The passing score for each section is 60 points out of 100. The HSK total score was computed by averaging across the two sections. Because scores on the HSK and in-class exam were highly correlated, we used the average of these two scores as the measurement for immediate attainment.

To assess long-term retention of Mandarin Chinese, students completed an additional HSK proficiency exam, administered online using a different sample test version. The link to the exam (hosted on Qualtrics; www.qualtrics.com) was sent out to all participants 90 days after the end of the course. Twenty-one out of 24 participants completed the long-term retention exam. Two of these participants' data were excluded from the analysis due to technical issues during the listening section of the online exam. Therefore 19 individuals (7 females and 12 males) had scores on their long-term retention exam.

### 2.3. Task fMRI experiment

In the fMRI scanner, participants discriminated the tone contours of pairs of real Mandarin words ("speech condition"), or pairs of nonspeech sinusoidal tones synthesized based on the words' natural pitch contours ("sinewave condition"). Participants were asked to press one of the two response buttons to indicate whether the pitch contours of consecutive pairs of sounds were the same (e.g., mǎi 买/yě 也) or different (e.g., tīng 听/xià 下) with respect to the lexical tone categories of Mandarin Chinese (high-level, rising, dipping, and falling).

The speech stimuli, consisting of 96 unique monosyllabic Mandarin words, were each recorded by two native speakers of Mandarin Chinese (1 male, 1 female), who were unknown to the participants. There were 24 stimuli for each of the four Mandarin lexical tones. All 96 monosyllabic words were chosen from the glossary of the Mandarin course textbook (see Procedure: Mandarin Training), which comprised 400 words. The stimulus set contained no homophonous word pairs and ranged from highly frequent words (e.g., "to have"/有) to less frequent words (e.g., "a quantifier for house"/套).

The sinusoidal tones were synthesized from the pitch contours of the natural speech stimuli. We extracted the pitch contour from the natural speech tokens using the "pitch-synchronous overlap and add" algorithm (Moulines and Charpentier, 1990) implemented in the software Praat (Boersma and Weenink, 2001). Three overtones of the fundamental frequency contour were added with a spectral tilt equal to that of the human voice (-6 dB/octave), to create a complex tone with rich harmonic structure that would better activate pitch-sensitive auditory cortices (Norman-Haignere et al., 2013). The low-pass amplitude contour of the corresponding natural speech token was then applied to the resynthesized complex tone. The sinewave stimuli were thus matched in pitch and intensity to the Mandarin speech, while lacking time-varying timbre (i.e., formants) of natural speech resulting from acoustic resonance of the vocal tract.

The task consisted of 192 tone-discrimination trials. Half of the trials contained sound pairs with the same tone (12 pairs × 4 lexical tone categories × 2 conditions), and the other half contained sound pairs with different tones (8 pairs × 6 different-tone combinations × 2 conditions). Each pair of speech sounds consisted of two Mandarin syllables that differed in consonants, vowels, and speakers (e.g., (1) guó 国spoken by a female voice, and (2) mā 妈spoken by a male voice). As a result, the stimuli always differed acoustically, both from each other and from listeners' previous in-class experiences. The acoustic-phonetic differences between stimuli meant that listeners could not simply detect whether the pitch contours of the two stimuli were identical. Instead, correct discrimination of the two stimuli always required participants to determine the phonological category of the tones. This was done to examine listeners' abstract, phonological processing of lexical tone, independent of acoustic-phonetic context. The pairs of sinewave sounds were generated from the corresponding female and male speech sounds respectively. Sinewave stimulus pairs were the non-speech analogs of the word pairs from each speech condition trial.

Participants completed two, 6-min runs in both pre- and post-training scan sessions. Each run consisted of six blocks of each of the two conditions (Mandarin speech and sinewave sounds), as well as six blocks of non-task resting baseline, in a pseudorandom order. Each block lasted 21.6s and contained either eight pairs of spoken Mandarin words, eight pairs of sinewave sounds, or silent fixation. Two blocks of the same condition were never immediately adjacent. The onset of the two sounds within each trial was 1s apart. After the sound presentation, participants had 1s to respond to the stimuli before the next trial started (ITI = 2.7s; Fig. 1).

Participants took part in two identical scanning sessions, the first before the 4-week Mandarin course and the second within one week after the end of the course. Prior to the first scanning session, participants practiced the task using 24 example pairs of sounds (12 speech pairs and 12 sinewave pairs) that were not included during the actual MRI session. Participants were informed of the various pitch contour categories (flat, rising, falling and dipping sounds) and instructed to decide if the two

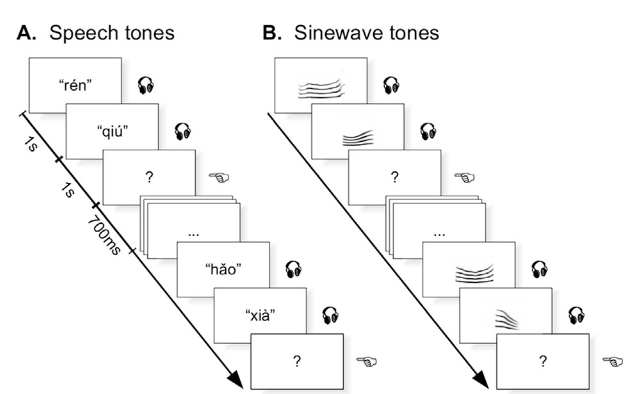


**Fig. 1.** Task design. Participants discriminated pairs of sounds in a blocked, sparse-sampling fMRI design. Pairs of audio stimuli were played during the 2-s silent delay between volume acquisitions, indicated here by headphones and example stimuli in either quoted pinyin for the Speech condition (A) or spectrograms for the Sinewave condition (B). Participants were prompted to respond via button press within a 1s interval after the auditory stimuli (indicated by "?"). A T2*-weighted, whole-brain volume was collected in the initial 0.5s of each TR.

sounds they heard had the same or different tones (or pitch contours). After each practice trial, participants were given automated feedback on the correctness of their response. Participants did not receive feedback during the actual in-scanner experiment.

Participants lay supine in the MRI scanner, held a response box with two response buttons in their right hand, and observed a projected computer display via an angled mirror suspended in front of their eyes. Auditory stimuli were presented binaurally using Sensimetric S14 ear-insert phones (Etymotic Research, Elk Grove Village, IL) at a comfortable listening level.

### 2.4. MRI data acquisition

Data were acquired on a Siemens Trio 3 T scanner with a 32-channel phased array head coil. Prior to the functional runs, a whole-head, high-resolution T1-weighted, multi-echo magnetization-prepared rapid gradient-echo (ME-MPRAGE) anatomical volume (van der Kouwe et al., 2008) aligned to the AC-PC plane was collected with the following acquisition parameters: TR = 2530 ms, TE = {1.64, 3.44, 5.24, 7.04 ms}, flip angle = 7.0°, TI = 1400 ms, voxel resolution = 1.0 $mm^3$, FOV = 220 × 220, 176 sagittal slices.

Two functional task runs containing 145 whole-brain volumes each were acquired using sparse-sampled T2*-weighted echo-planar imaging with simultaneous multi-slice scans acquisition (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012) with the following acquisition parameters: TR = 2700 ms, TA (acquisition time) = 500 ms, TE = 30 ms, flip angle = 75°, voxel resolution = 3.0 $mm^3$, FOV = 210 × 210 mm, multi-band acceleration factor = 6, and 36 transverse slices parallel to the AC-PC plane acquired in an interleaved order from superior to inferior. Sparse-sampling (Hall et al., 1999) was used to allow auditory stimuli to be presented in silence. One participant did not complete the second scan session and therefore was excluded from analyses involving the post fMRI data.

Resting state functional MRI scans lasted 6.25 min, during which participants were instructed to “keep your eyes open and think of nothing in particular.” Resting scans consisted of 37 interleaved oblique, 3.5 mm-thick axial slices, covering the entire brain (TR = 2500 ms, TE = 30 ms, flip angle = 90°, FOV 224 × 224 mm, with 3.5 mm isotropic voxels). Prospective motion correction (PACE; Thesen et al., 2000) was used to mitigate artifacts due to head motion. One other participant did not complete the second scan session and therefore was excluded from fMRI analyses involving the post-training scan.

### 2.5. In-scanner behavior analysis

Trials with no response, or where participants did not respond before the beginning of the next trial, were excluded from the analysis (mean: 11.6%; SD: 12.9% out of all trials). Tone sensitivity ($d'$) was calculated using the “hit” and “false alarm” rates for each condition for each participant in both pre- and post-training sessions. Data were analyzed by a mixed effects linear model (lme4 v1.1.6; Bates et al., 2015) using R v3.2.4 (R Core Team, 2015), with sessions (pre vs. post) and conditions (speech vs. sinewave) as fixed effects. The final model, deduced from the maximal model using backward model comparison, included random intercepts for participants and random slopes for sessions and conditions (Barr et al., 2013).

### 2.6. Task fMRI data analysis

Cortical surface reconstruction and parcellation of the T1-weighted anatomical images were performed using the default processing stream FreeSurfer v5.3.0 (Dale et al., 1999). Functional data were analyzed in FSL v5.0.6 (Smith et al., 2004) using Nipype v0.8 workflows (Gorgolewski et al., 2011) implemented via Brain Imaging Pipelines (BIPS; Ghosh et al., 2012). The 4D functional timeseries from each functional run was first realigned to its first volume using normalized correlation optimization and cubic spline interpolation. Motion and intensity outliers (functional volumes exceeding 1 mm in total differential motion relative to the previous volume or differing from the mean image intensity by > 3 SD) were identified and regressed out of the hypothesized timeseries (Siegel et al., 2014). On average, this procedure discarded 2.9% of participants' functional volumes. The composite motion measure was calculated by converting the 3 rotation and 3 translation parameters into vectors reflecting the trajectories of 6 points located on the center of each of the faces of a bounding box around the brain. We then used the maximum scan-to-scan movement of any of these points as the composite score for each time point, and then averaged across time points to calculate a mean composite motion score for the entire scan for each individual participant. There was no difference between pre- and post-training sessions in terms of motion, measured by the average of the six motion parameters across trials ($M_{pre} = .09$ mm, $SD_{pre} = .03$ mm; $M_{post} = .10$ mm, $SD_{post} = .04$ mm; $t$ (22) = 1.15, $p = .26$). Motion parameters were not associated with Mandarin learning outcomes ($p$'s > 0.22). Uncorrelated timeseries noise was attenuated by spatially low-pass filtering with a 3D Gaussian kernel of 6 mm FWHM.

Model design included two task regressors (“Speech” and “Sinewave” conditions), six motion parameters, and individual regressors for any outlier volumes. Vectors for task regressors were determined by convolving a vector of event onset times with their durations, convolving the resulting vector with a canonical HRF (gamma difference), and resampling the resulting time series to include only timepoints when scanner data were actually acquired (Perrachione and Ghosh, 2013). The key contrast for all the analyses in this study is the difference between the speech and sinewave conditions, which dissociates speech and linguistic processing from acoustic processing of the lexical tones. Within-subject estimation of the general linear model and contrasts were conducted for each run in participants' native EPI space.

The coregistration transformation between the mean functional volume of each run and the participant's T1-weighted anatomical image was calculated using the six degree of freedom rigid-body alignment via boundary-based optimization implemented in the FreeSurfer program *BBRegister* (Greve and Fischl, 2009). Participants' high-resolution structural images were normalized to the MNI152 standard template via nonlinear symmetric diffeomorphic mapping implemented in ANTS v1.9 (Avants et al., 2011). Each participant's transformation matrix and deformation field from this spatial normalization were applied to their coregistered first-level functional images to align them to the common space. Data in normalized space from participants' two functional runs were combined using a fixed-effects model in FSL. Group-level statistical parametric (“activation”) maps were computed by estimating a mixed-effects general linear model (Woolrich et al., 2009) in FSL, with contrasts of interest including within-group means and between-session differences for each of the within-participant model contrasts described above. Group-level statistics were based on non-parametric permutation method (FSL 5.0.9's *Randomise* function, 5000 permutations), recommended for controlling type I error rate (Silver et al., 2011; Eklund et al., 2016). The resulting group activation maps were thresholded at voxel-level cluster determining threshold $p < .005$ (two-sided), and correction for multiple comparisons was accomplished by controlling cluster-level family-wise error (FWE) at $p < .05$.

In correlation analyses, participants' in-scanner lexical tone discrimination accuracy was entered as a regressor in the mixed-effects models to test the relations between fMRI signal intensities and behavioral parameters. A small volume correction (SVC) for multiple tests was applied using the same nonparametric permutation method as above. SVC-volumes were defined by the bilateral IFG regions using the Jülich probabilistic atlas (2 mm atlas thresholded at 50% probability: Eickhoff et al., 2007), which results in a 24.6-$cm^3$ mask used for SVC. Clusters survived voxel-level threshold $p < .01$ (two-sided) and FWE cluster corrected $p < .05$.

### 2.7. Resting-state functional connectivity data analysis

SPM8 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm) was used to preprocess resting-state functional MRI data. Functional images were slice-time and motion corrected, registered to structural scans, normalized to an MNI template, and smoothed with a 6 mm FWHM Gaussian kernel.

Prior to spatial filtering, an artifact detection algorithm (ART; http://www.nitrc.org/projects/artifact_detect), was used to identify outlier data points, defined as volumes that exceeded 3 standard deviations from mean global brain activation across the entire volume, or a composite movement threshold of 1 mm framewise displacement. The mean composite movement measure was calculated for each individual participant in the same way as described in the task fMRI. Because movement during resting state scans can artificially induce short-range and mask long-range correlations (Power et al., 2012), we also determined that the pre- and post-training sessions did not differ significantly from each other in the amount of composite motion ($M_{pre} = .07$ mm, $SD_{pre} = .04$ mm; $M_{post} = .09$ mm, $SD_{post} = .05$ mm; paired $t$ (22) = 1.45, $p = .16$). These motion parameters were not associated with Mandarin learning outcomes ($p$'s > 0.11).

We used a seed-driven approach with custom software developed in Matlab, "Conn toolbox," (http://www.nitrc.org/projects/conn/), for functional connectivity analyses (Whitfield-Gabrieli and Nieto-Castanon, 2012). Resting-state functional data were temporally band-pass filtered (0.009 Hz–0.08 Hz) in order to remove signals of non-neurophysiological origin. Head motion parameters (3 rotation, 3 translation), as well as their first derivatives, and outlier nuisance regressors (one for each motion outlier consisting of all zeros and a 1 at the respective outlier time point) were regressed out of the first level GLMs (Siegel et al., 2014). Spurious sources of non-neurophysiological noise were accounted for through an anatomical component based (aCompCor) approach (Behzadi et al., 2007). Each participant's structural image was segmented into white matter (WM), grey matter (GM), and cerebral spinal fluid (CSF) using SPM8. WM and CSF masks were eroded by one voxel to avoid partial volume effects with adjacent grey matter. The first 3 principal components of the signals from the eroded WM and CSF noise ROIs were removed with regression. Pearson's correlation coefficients were calculated between the seed time-series and the time course of all other voxels. The resulting Pearson's correlation coefficients were z-transformed using the Fisher transformation to conform to the assumptions of the General Linear Model (normally distributed) for second-level analyses.

We tested whether more successful learning was related to the magnitude of change in functional connectivity between the regions that were functionally activated during tone discrimination task after training. We related the connectivity change from pre-training to post-training sessions among the eight functional ROIs (bilateral superior temporal gyrus (STG), bilateral superior parietal lobule (SPL), bilateral IFG, and bilateral inferior temporal gyrus (ITG)) to individuals' Mandarin attainment. All ROIs were defined by the significant functional clusters from the post-training scan (Fig. 3B) and their right-hemisphere counterparts. Twenty-eight ROI-to-ROI connections were examined simultaneously and correction for multiple comparisons was accomplished by controlling seed-level false discovery rate (FDR) at $p < .05$. In order to test whether spontaneous brain activation patterns are related to the neural organization during speech perception, we also examined the relationship between the plasticity in resting-state connectivity and the plasticity in task fMRI.

### 2.8. Predicting Mandarin long-term retention

Because it is every language learner's ultimate goal to retain newly learned language skills for the long term after training, we aimed to identify the behavioral and neural features that reliably predicted long-term retention of out-of-sample individuals using cross-validation in the 19 learners (7 female, 12 male) for whom scores on the long-term retention exam were available.

Across 1000 permutations, we randomly chose four out of the nineteen participants as a testing set and then performed linear regression in the remaining participants with the prediction model. The root mean square error (RMSE) was computed across all testing participants to measure the deviance of prediction error. As a result, the prediction model generates between 150 and 250 predicted scores for each participant and 1000 RMSE values across all resamplings. The performance of a prediction model is reported in two measures: 1) the R-Squared value, measuring the extent of correlation between each learner's median predicted score and the actual score, and 2) the median RMSE among 1000 resamplings. The significance of each model was tested by comparing the linear regression model between the median predicted score and the actual score with the null hypothesis that there was no relationship between the predicted and the actual scores. The cross-validation analysis was conducted with the *caret* package (Kuhn, 2008) in R 3.2.4 (R Core Team, 2015).

Model comparison was conducted by comparing the RMSE distribution among different predictor types (behavioral model, neural model, and behavioral + neural model). Pairwise comparison between predictor types was achieved by simultaneous tests of general linear hypotheses (multcomp 1.4–6; Hothorn et al., 2008). P-values were adjusted by the Tukey method for multiple comparisons.

## 3. Results

### 3.1. Behavioral and neural plasticity in successful classroom-based Mandarin learning

We combined both task and resting-state fMRI to examine the neural changes induced by one month of classroom Mandarin training. Following Wong et al. (2007), we measured neural responses to speech versus sinewave tones before and after Mandarin training to investigate neural plasticity of foreign speech perception. By using resting-state fMRI, we investigated change in spontaneous connectivity among functionally activated region during speech perception.

### 3.2. Improvement in lexical tone discrimination and Mandarin proficiency

Participants' performance in both the speech and the sinewave conditions in both the pre- and post-training sessions was significantly above 50% chance level (Fig. 2A; pre-training speech accuracy: 58%, SD = 12%, $t(23) = 3.75$, $p = .001$; pre-training sinewave accuracy: 66%, SD = 16%, $t(23) = 7.27$, $p < .001$; post-training speech accuracy: 65%, SD = 10%, $t$

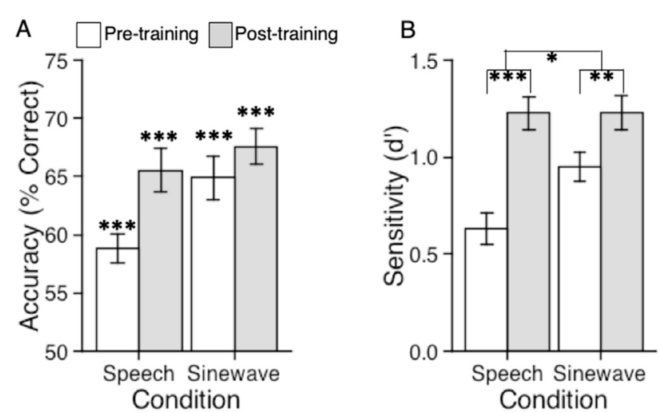


**Fig. 2.** In-scanner performance on the tone-discrimination task. Before Mandarin training (light bars), participants responded more accurately (A) and had higher $d'$ values (B) for Sinewave stimuli than for Speech. After Mandarin training (dark bars), participants' tone discrimination accuracy (A) and $d'$ values (B) improved more for Speech than Sinewave stimuli. Error bars are within-subject standard error of the mean (Morey, 2008). *: $p < .05$; ***: $p < .001$.

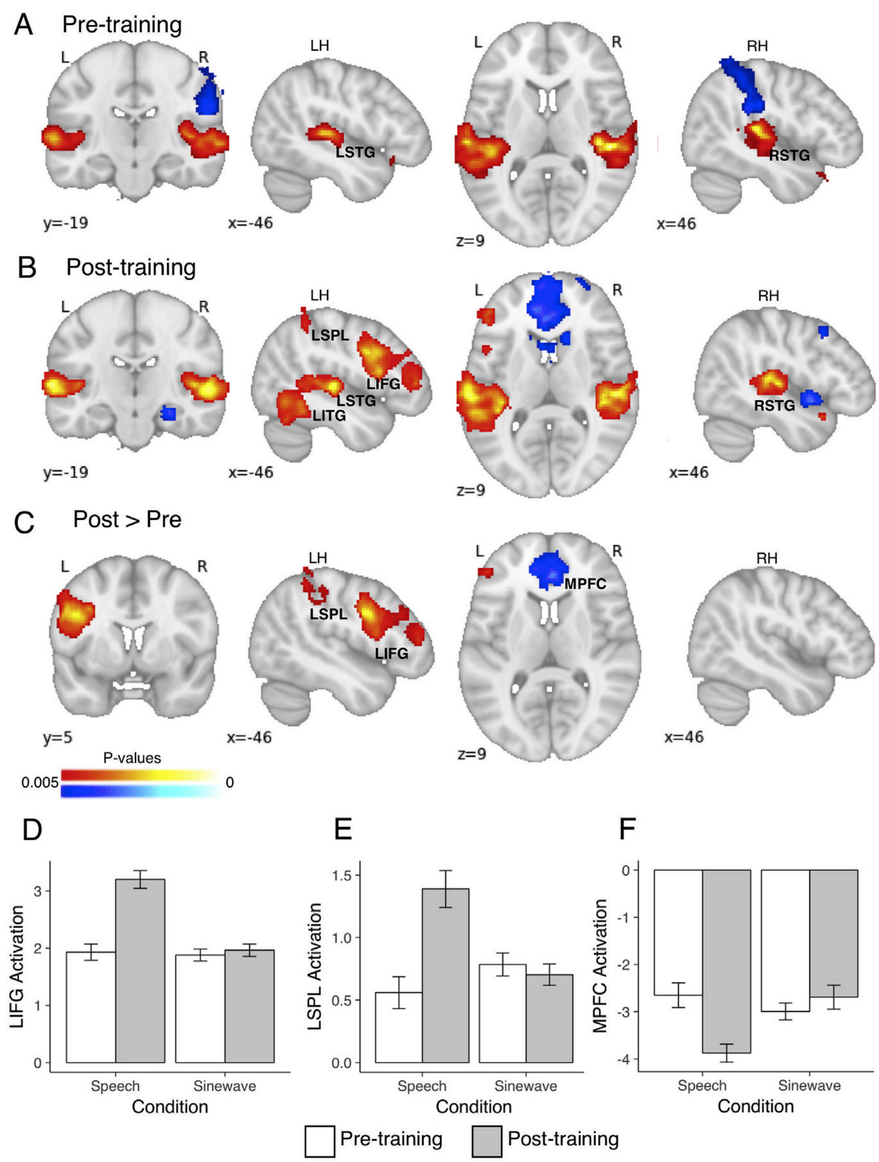


**Fig. 3.** Training-induced changes in task-based fMRI. (A-C) Speech vs. Sinewave effect in the whole-brain analysis at voxelwise $p < .005$, cluster-level FWE-corrected $p < .05$. (A) Speech vs. Sinewave in the pre-training session, (B) post-training session, and (C) interaction between the Speech vs. Sinewave condition effect and the Post vs. Pre-training session effect. (D-F) Bar graphs illustrate the interactions in the three regions of interest labelled in (C): (D) left inferior frontal gyrus (LIFG), (E) left superior parietal lobule (LSPL), and (F) medial prefrontal cortex (MPFC). Light grey bars indicate BOLD response (beta weight) before training; dark grey bars indicate BOLD response (beta weight) after training. Error bars are within-subject standard error of the mean.

$(23) = 4.76$, $p < .001$; post-training sinewave accuracy: 68%, SD = 12%, $t(23) = 4.76$, $p < .001$).

Discriminability of sinewave tones ($d'$) was greater than that of speech tones across both sessions (main effect of condition; $\beta = 0.32$, *s. e.* $= 0.12$, $t = 2.62$, $p = .012$) (Fig. 2B). After training, participants' $d'$ improved significantly overall (main effect of session; $\beta = 0.59$, *s. e.* $= 0.10$, $t = 5.91$, $p < .001$). The extent of improvement in the speech condition was significantly greater than that in the sinewave condition (condition × session interaction; $\beta = 0.32$, *s. e.* $= 0.14$, $t = 2.27$, $p = .028$). Post-hoc pairwise comparisons showed that training-induced improvement of $d'$ was nonetheless significant in both the speech condition (paired $t(23) = 5.93$, $p < .001$) and the sinewave condition (paired $t(23) = 2.72$, $p = .01$).

On the last day of the Mandarin course, participants' performance on the regular final exam was significantly correlated with their performance on the HSK ($r = 0.68$, $p < .001$). Twenty-two out of 24 participants passed the HSK. Attainment, measured by the average of these two exams, ranged from 52.9 to 91.6 with a mean of 73.7 (out of 100). Ninety days after the end of the course, participants' long-term retention, measured by a second version of the HSK, ranged from 32.5 to 95.0 with a mean of 54.7 (out of 100), which was significantly lower than their earlier HSK scores (paired two-sided $t(18) = 3.39$, $p = .003$).

### 3.3. Group-level increase in neural sensitivity to lexical tones at left SPL and left IFG

The fMRI responses elicited by Mandarin speech tones before and after training are illustrated in a functional contrast map ($z$-statistics) comparing speech to sinewave tones (Fig. 3, Table 1). Before language training, participants showed significantly greater activation in response to speech tones than to sinewave tones in bilateral superior temporal gyri (STG) (Fig. 3A). After language training, in addition to bilateral STG, participants showed greater activation in response to speech than sinewave tones in left IFG, left SPL and left ITG (Fig. 3B). Comparing pre-training and post-training sessions, left IFG and left SPL showed significantly greater increase of activation to speech tones, compared to sinewave tones (Fig. 3C–E).

In addition, language training resulted in greater deactivation in the medial prefrontal cortex (MPFC), an integral node of default mode network areas (DMN; Raichle et al., 2001; Buckner et al., 2008), in response to the speech than sinewave tones. This interaction between session and condition was mainly driven by more deactivation of MPFC in response to speech tones, rather than changes in the response to sinewave tones (Fig. 3C and F).

These patterns of organization and reorganization in the neural

**Table 1**
**Regions of significant clusters in the tone-discrimination task.** Regions labelled according to Harvard-Oxford Cortical and Subcortical Structural Atlases. Abbreviations: angular gyrus (AG); inferior frontal gyrus (IFG); inferior temporal gyrus (ITG); medial prefrontal cortex (MPFC); middle frontal gyrus (MFG); Precentral gyrus (PreCG); Postcentral gyrus (PostCG); superior temporal gyrus (STG); superior parietal lobule (SPL); supramarginal gyrus (SMG).

| Cortical Regions | Peak voxel | | | | Cluster Volume |
|---|---|---|---|---|---|
| | X | Y | Z | $Z_{max}$ | (2 × 2 × 2 voxel size) |
| **Main effect of condition** | | | | | |
| Speech > Sinewave | | | | | |
| Right STG | 44 | −24 | 6 | 13 | 3548 voxels |
| Left STG | −64 | −24 | 4 | 11.9 | 3809 voxels |
| Sinewave > Speech | | | | | |
| Right SPL/SMG | 58 | −24 | 32 | 8.55 | 4770 voxels |
| **Main effect of session** | | | | | |
| Pre-training > Post-training | – | – | – | – | – |
| Post-training > Pre-training | – | – | – | – | – |
| **Interaction** | | | | | |
| Session x Condition | | | | | |
| Left IFG/PreCG | −44 | 0 | 32 | 11.6 | 2436 voxels |
| Left SPL/SMG/AG | −28 | −60 | 50 | 6.39 | 2204 voxels |
| MPFC | 4 | 40 | 6 | 7.03 | 1735 voxels |
| **Within each session** | | | | | |
| Pre-training | | | | | |
| Speech > Sinewave | | | | | |
| Right STG | 50 | −26 | 8 | 13.3 | 2784 voxels |
| Left STG | −64 | −24 | 4 | 11.0 | 2721 voxels |
| Sinewave > Speech | | | | | |
| Right SPL/SMG/PostCG | 58 | −26 | 36 | 7.86 | 3570 voxels |
| Post-training | | | | | |
| Speech > Sinewave | | | | | |
| Left STG/ITG | −58 | −28 | 4 | 10.4 | 4002 voxels |
| Left IFG/MFG/PreCG | −40 | 6 | 32 | 9.48 | 2682 voxels |
| Right STG | 62 | −12 | −2 | 9.96 | 2281 voxels |
| Left SPL/SMG | −24 | −70 | 44 | 6.76 | 1639 voxels |
| Sinewave > Speech | | | | | |
| MPFC | −10 | 48 | 16 | 6.69 | 8427 voxels |

sensitivity to lexical tone processing remained the same in additional analyses in which we controlled for participants' in-scanner behavioral accuracy in the group-level models, and in which we regressed out trials with incorrect responses in the first-level models. There was no significant difference in the patterns of activation between correct and incorrect trials either before or after training. Supplemental analyses within only the 19 monolingual learners showed the same pattern of plasticity (Supplementary Results).

### *3.4. Association between successful language learning and the plasticity of inter-hemispheric resting-state functional connectivity*

We chose the functionally activated regions during the post-training tone-discrimination task and their right hemisphere homologues as ROIs for resting-state functional connectivity analyses, resulting in eight functional seeds (left and right STG, SPL, IFG and ITG). After one month of training, learners did not exhibit significant increase of connectivity among these regions on average. Importantly, however, greater pre-post increases in RIFG-LSPL ($t\,(21) = 2.82$, $p = .010$, seed-level FDR-corrected $p < .05$) and RIFG-LIFG ($t\,(21) = 3.16$, $p = .005$, seed-level FDR-corrected $p < .05$) connectivity were significantly associated with more successful Mandarin attainment (Fig. 4A–C) and better post-training speech tone-discrimination accuracy (Table 2).

### *3.5. Association between successful language learning and neural sensitivity to lexical tones in right IFG*

Left IFG showed group-level increase of sensitivity to speech tone discrimination after training (Fig. 3C), while right IFG appears to be the connectivity hub that was important for both speech perception and successful language learning. Therefore, we further examined the functional role of right IFG, as well as left IFG, by relating the neural responses to speech tones before and after training to individuals' in-scanner performance. Before Mandarin learning, greater activation in response to Mandarin speech tones in right IFG was associated with better lexical tone discrimination accuracy (Fig. 5A). In contrast, after Mandarin learning, greater activation to lexical tones in right IFG was associated with worse post-training lexical tone discrimination accuracy (Fig. 5B). Learners' pre-training right IFG activation, extracted from the significant cluster in Fig. 5A, was negatively correlated with post-training right IFG activation, extracted from the significant cluster in Fig. 5B (Spearman's *rho* = −0.55, $p = .007$, Fig. 5C). Greater pre-post reduction of right IFG activation was significantly associated with better post-training tone discrimination (Spearman's *rho* = −0.78, $p < .0001$; Fig. 5C) and better

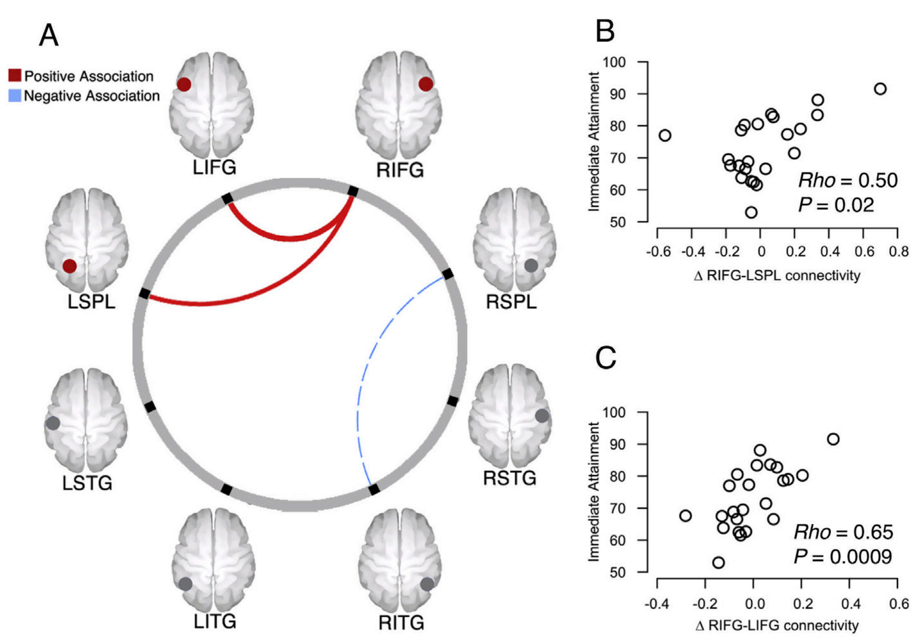


**Fig. 4.** The pre-to-post change of resting-state connectivity among bilateral STG, bilateral SPL, bilateral IFG and bilateral ITG, functionally defined by the significantly activated clusters during the post-training tone-discrimination task and their right-hemisphere homologues. (A) Relationship between the change of ROI-to-ROI connections and Mandarin attainment. Red connections indicate positive association and blue connections indicate negative association. The line thickness is proportional to the effect size. The solid connection lines show significant effects after correction for multiple comparisons (seed-level FDR-corrected $p < .05$). The dash connection line shows the uncorrected significant effect ($p < .05$), which was no longer significant after correction for multiple comparisons. (B-C) Greater increases in RIFG-LSPL connectivity (B) and RIFG-LIFG connectivity (C) were associated with more successful Mandarin attainment.

**Table 2**
Correlations among neural and behavioral plasticity measures.

| | fMRI activation[a] change | | | rs-fMRI connectivity change | | Behavioral change |
|---|---|---|---|---|---|---|
| | LIFG | LSPL | RIFG | RIFG-LIFG | RIFG-LSPL | Tone discrimination |
| **fMRI activation[a] change** | | | | | | |
| LSPL | .74* (.0001) | | | | | |
| RIFG | .27 (.21) | 32 (.13) | | | | |
| **rs-fMRI connectivity change** | | | | | | |
| RIFG-LIFG | .14 (.53) | .05 (.82) | -.69* (.004) | | | |
| RIFG-LSPL | .14 (.54) | -.00 (.98) | -.58* (.005) | .50* (.01) | | |
| **Behavioral change** | | | | | | |
| Tone discrimination | -.06 (.78) | -.09 (.68) | -.78* (<.0001) | .65* (.0009) | .62* (.001) | |
| Immediate attainment | .16 (.46) | .21 (.35) | -.60* (.003 | .62* (.002) | .50* (.02) | .57* (.004) |

Spearman's rho and significance (uncorrected *p*) are shown for all bivariate correlations.
* FDR-corrected $p < .05$.
[a] fMRI activation for speech > sinewave contrast.

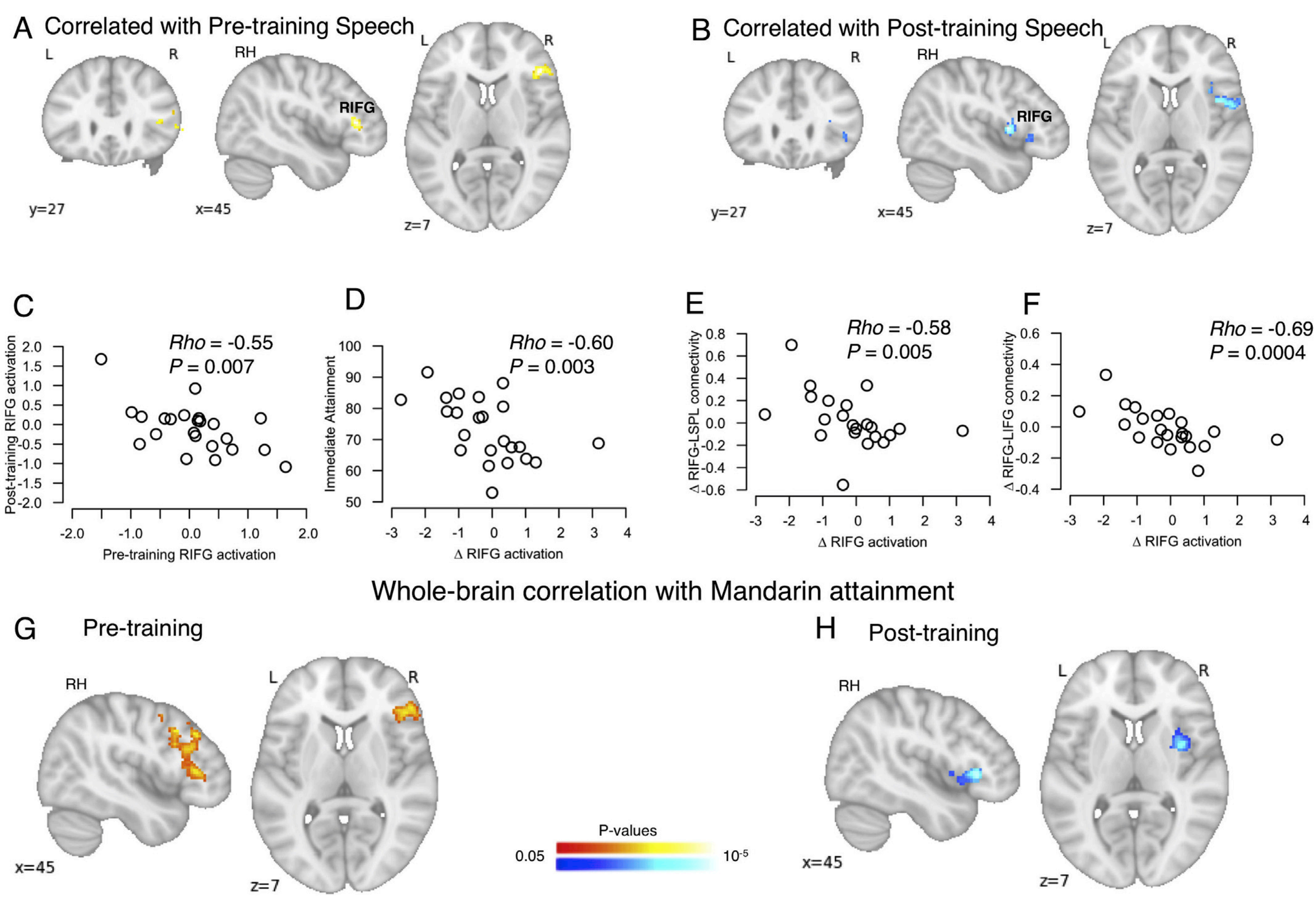


**Fig. 5.** Correlations between fMRI response (Speech vs. Sinewave) and learners' behavior. (A) Before training, greater activation in right inferior frontal gyrus (RIFG) was associated with more accurate pre-training, in-scanner speech tone discrimination. (B) After training, less activation in RIFG was associated with more accurate post-training, in-scanner speech tone discrimination. (In panels A and B, significant clusters are shown at voxelwise $p < .01$ two-sided, cluster-level FWE-corrected $p < .05$). (C) Greater pre-training RIFG activation was associated with less post-training RIFG fMRI activation. (D-F) Greater reduction of RIFG fMRI activation over the course of training was associated with better immediate Mandarin attainment (D), greater enhancement in RIFG-LSPL connectivity (E), and greater enhancement in RIFG-LIFG connectivity (F). (G & H) Whole-brain analyses of the association between fMRI response and Mandarin attainment (significant clusters are shown at voxelwise $p < .05$ two-sided, cluster-level FWE-corrected $p < .05$). (G) Greater pre-training right IFG activation is associated with better Mandarin attainment. (H) Less pre-training right IFG and right insula activation is associated with better Mandarin attainment.

Mandarin attainment (Spearman's *rho* = −0.60, *p* = .003; Fig. 5D & Table 2). Moreover, functional disengagement of right IFG was also significantly associated with greater enhancement in resting-state connectivity between RIFG and LSPL (Spearman's *rho* = −0.58, *p* = .005; Fig. 5E) as well as between RIFG and LIFG (Spearman's *rho* = −0.69, *p* = .0004; Fig. 5F & Table 2). No voxels at the left IFG were significantly related to learners' in-scanner behavior. These patterns remained the same when we included only the 19 monolingual learners (Supplementary Table 2), and when we accounted for variations in IQ (Supplementary Results).

Exploratory whole-brain analyses on the brain-proficiency association verified that right IFG was positively associated with Mandarin attainment before training and negatively associated with Mandarin attainment after training. No region in the left hemisphere was associated with Mandarin attainment either before or after training (Fig. 5Gand H).

Although left IFG and the left SPL showed group-level increase of activation after training (Fig. 3C), the magnitudes of post *vs.* pre differences in these two regions were neither associated with the other functional plasticity measures (right IFG activation and inter-hemispheric connectivity) nor associated with learners' behavior (in-scanner performance and Mandarin attainment) (Table 2).

We also considered whether the functional organization of right IFG was specifically associated with aptitude for learning Mandarin or other tonal languages. Follow-up analyses verified that participants' MLAT scores (which are measured with non-tonal language stimuli on phonetic coding, grammatical, rote learning and rule induction) were also significantly correlated with pre-training right IFG engagement, right IFG disengagement after training, as well as interhemispheric connectivity enhancement involving right IFG after training (Supplementary Results).

### 3.6. Neural preparedness and plasticity of right IFG predict long-term retention of Mandarin

We examined whether fMRI activation prior to training (neural preparedness) and fMRI activation change over the course of learning (neural plasticity) in right IFG could predict out-of-sample individuals' long-term retention, measured 90 days after the course. Given the strong association between the fMRI activation in right IFG and individual differences in both pre-training tone discrimination and Mandarin attainment (Fig. 5), we asked whether the neural predictors based on right IFG fMRI activation perform better than these behavioral measures in predicting long-term retention. The performance of these cross-validated prediction models is shown in Table 3.

### 3.7. Neural preparedness at right IFG predicts long-term retention

We used behavioral and neural sensitivity to lexical tones prior to Mandarin training as indices of individuals' language-learning preparedness, and assessed the ability of these measures to predict out-of-sample individuals' long-term retention. Behavioral preparedness was measured by pre-training speech tone discrimination accuracy. Neural preparedness was measured as pre-training fMRI activation (speech *vs.* sinewave) extracted from the right IFG cluster that showed a significant correlation with in-scanner behavior (Fig. 5A).

The behavioral preparedness model did not significantly predict long-term retention ($F$ (1,17) = 0.18, $p$ = .68) and accounted for only 1% of the variance in long-term retention scores. In contrast, the neural preparedness model significantly predicted long-term retention ($F$ (1,17) = 4.84, $p$ = .04) and accounted for 22% of the variance (Table 3). The model combining the neural and behavioral preparedness predictors did not significantly predict long-term retention ($F$ (1,17) = 2.53, $p$ = .13) and accounted for 13% of the variance. Model comparison based on the distribution of prediction error (RMSE) showed that the neural preparedness model was a significantly better model compared to the behavioral preparedness model ($z$ = −19.06, $p$ < .001). The combined model was not significantly better than the neural preparedness model ($p$ > .5, Table 3).

**Table 3**
Performance of models predicting out-of-sample long-term retention.

| Model features | Preparedness models[a] | | Plasticity models[b] | |
|---|---|---|---|---|
| | $R^2$ | Median RMSE | $R^2$ | Median RMSE |
| Behavioral | .01 | 18.06 | .09 | 17.69 |
| Neural | .22[c] | 16.71 | .22[c] | 16.68 |
| Behavioral + Neural | .08 | 18.00 | .21[c] | 17.20 |

[a] Preparedness models include pre-training tone discrimination accuracy (behavioral) and pre-training right IFG activation to Mandarin speech (neural).
[b] Plasticity models include Mandarin immediate attainment (behavioral) and pre-to-post changes in right IFG activation to Mandarin speech (neural).
[c] Model significantly predicts out-of-sample long-term retention.

### 3.8. Neural plasticity in right IFG predicts long-term retention

Behavioral plasticity was operationalized as learners' attainment of holistic Mandarin proficiency at the conclusion of the course. Neural plasticity was operationalized as the reduction of fMRI activation to speech tones in right IFG. We constructed prediction models of long-term Mandarin retention based on behavioral plasticity, neural plasticity, and their combination.

The behavioral plasticity model did not significantly predict long-term retention ($F$ (1,17) = 1.17, $p$ = .210) and only accounted for 9% of the variance in long-term retention scores. In contrast, the neural plasticity model significantly predicted long-term retention ($F$ (1,17) = 4.73, $p$ = .044) and accounted for 22% of the variance. The model combining the neural and behavioral plasticity predictors also significantly predicted long-term retention ($F$ (1,17) = 4.47, $p$ = .049) and accounted for 21% of the variance (Table 3). Model comparison based on the distribution of prediction error (RMSE) showed that the neural plasticity model was a significantly better model compared to the behavioral plasticity model (z = −6.47, $p$ < .001). The combined model was not significantly better than the neural plasticity model ($p$ > .5) (Table 3).

## 4. Discussion

The present study revealed that individual differences in right IFG engagement in speech perception, both before and after intensive language learning, were strongly related to learners' success at holistic, high-level acquisition of Mandarin Chinese. Learning Mandarin Chinese resulted in increased activation in left IFG and left SPL in response to Mandarin speech sounds, but variation in neither left IFG nor left SPL activation across individuals was related to language learning outcomes. Instead, individual differences in both neural preparedness (initial right IFG engagement) and neural plasticity (subsequent right IFG disengagement) predicted long-term language learning success. Disengagement of right IFG during Mandarin speech perception occurred in conjunction with increased resting-state connectivity between left and right hemispheres after intensive classroom learning. The increased functional connectivity may reflect plasticity in the communication between hemispheres that facilitated the growing disengagement of right IFG in speech processing.

### 4.1. Importance of neural sensitivity to lexical tones in right IFG prior to learning

Greater pre-training activation to lexical tones in right IFG was associated with better lexical tone discrimination accuracy, better overall Mandarin attainment immediately following the instructional period, and better long-term retention of holistic Mandarin skills 90 days after training. Right IFG has previously been shown to be sensitive principally to acoustic, rather than phonological differences in speech sounds (Wong et al., 2004; Zatorre and Gandour, 2008; Myers et al., 2009). The present results expand our understanding of the role of processing speech acoustics in foreign language learning, revealing that this ability is both more foundational and has longer-lasting effects than previously thought. Greater right IFG sensitivity to speech acoustic differences prior to learning may prepare learners for accurate speech sound discrimination during the initial stage of learning, which in turn facilitates acquisition and retention of higher-level linguistic skills in the long term.

These fMRI results also provide new evidence for the specific functional role played by structural predictors of foreign language learning in the right hemisphere. In the same group of learners, pre-training white-matter microstructure in the right, but not the left, hemisphere was significantly associated with individual differences in Mandarin attainment (Qi et al., 2015; cf. Wong et al., 2011). In particular, greater fractional anisotropy of right superior longitudinal fasciculus and right inferior longitudinal fasciculus – both reflecting structural connectivity of right IFG – correlated with better holistic Mandarin attainment. The current findings provide converging functional evidence that pre-training right IFG is actively engaged in initial Mandarin learning.

The observed role of the right hemisphere is unlikely to result from the unique phonological feature of lexical tones in languages like Mandarin, as individual differences in phonetic learning of Hindi consonant have also been related to variation in frontal-temporal landmarks in the right hemisphere (Golestani et al., 2007). The significant correlation between right IFG and MLAT scores in our sample, a measure of foreign language learning aptitude using non-tonal testing materials, implicates a language-general account for the role of the right hemisphere. An important charge for future work will be to determine how early perceptual processing in the right hemisphere also affects the holistic acquisition and retention of non-tonal language skills (cf. Hoeft et al., 2011; Myers and Swan, 2012).

### 4.2. Importance of functional transition from right to left IFG over the course of learning

Left IFG and left SPL are actively engaged in native speech sound processing (Burton, 2001; Hsieh et al., 2001; Gandour et al., 2002, 2004; Blumstein et al., 2005; Díaz et al., 2008; Myers et al., 2009). Learners of non-native speech sounds show increased sensitivity in these regions after intensive training (Golestani and Zatorre, 2004; Myers and Swan, 2012). While we replicated findings from these laboratory-based training studies, the present study also provides critical new insight into the functional role of the right hemisphere in language learning. Successful acquisition and retention of a foreign language hinged principally on disengagement of right IFG function, rather than engagement of left frontal and parietal regions. The importance of greater disengagement of the right IFG after training reveals that while learning aptitude may be a function of the initial state of the system (as greater initial engagement of right IFG was also associated with better learning), successful learning also depends on the extent to which neural systems are able to re-organize during learning.

After completing the Mandarin course, better speech perception ability and greater Mandarin attainment and retention no longer related to *greater* right IFG activation, but instead *lesser* right IFG engagement. Previous research has shown that explicit training on non-native consonant identification appears to increase both left and right IFG responses to newly learned speech sounds, suggesting strengthened acoustic analysis at a group level (Callan et al., 2003; Wang et al., 2003; Golestani and Zatorre, 2004). Paradoxically, learning success of these non-native contrasts was associated with less increase of activation in bilateral IFG, which was interpreted as more automated, bottom-up, and efficient processing (Golestani and Zatorre, 2004). When the scope of training, however, goes beyond just phonetics to the whole, integrated language, our results suggest an apparent engagement transitioning to the left-hemisphere network. It is possible that as the phonological analysis becomes necessary for higher-level language processing, the acoustic analysis of speech sounds supported by the right hemisphere might no longer be optimal (Blumstein et al., 2005; Hickok and Poeppel, 2007; Fedorenko et al., 2011; Friederici, 2011; Myers, 2014). Our observations of the importance of right-IFG disengagement after learning may be analogous to neuromodulation studies where suppression of right IFG function facilitated naming and word learning in both healthy adults and those with aphasia (Naeser et al., 2011, 2012; Nicolo et al., 2016).

The finding that successful learners had increased resting-state connectivity between right IFG and the left hemisphere language regions after training is suggestive of a cross-hemispheric mechanism that might subserve or reflect the transition between right and left-hemisphere engagement (and perhaps consequently right-IFG disengagement). This plasticity in functional connectivity parallels plasticity in structural connectivity: Language learners displayed progressive strengthening of white-matter structural connectivity in the frontal tracts connecting left language areas and their right homologues over the course of 9-month Mandarin learning (Schlegel et al., 2012). Given the high degrees of hemispheric symmetry in most resting-state networks (Nir et al., 2008; Smith et al., 2012), these findings provide additional evidence for a model in which cross-hemispheric connectivity serves as means for contralateral inhibition and is necessary for left lateralization for language functions (Cook, 1984; Bitan et al., 2010; Hinkley et al., 2016). Therefore, stronger resting-state connectivity between hemispheres might facilitate more effective left-to-right inhibition that leads to greater down-regulation of the right hemisphere and potentially less functional competition.

The observed growth of activation in left IFG and left SPL is consistent with its role in language learning and language processing, but there was no relation between variation in that growth and learners' holistic learning success. Individual differences in left IFG functional plasticity have been associated with successful speech-sound learning (Ventura-Campos et al., 2013) after laboratory-based phonological training. One structural MRI study found that greater growth of left IFG grey matter volume was associated with greater German proficiency improvement measured after 5 months of immersive exposure to German (Stein et al., 2012). These structural findings suggest that variation in plasticity in left IFG may be related to later stages of whole-language learning than were examined in the present 1-month study. Further studies will be needed to develop a comprehensive neurobiological model of adult language learning that integrates structural and functional measures from initial learning through long-term mastery.

In sum, this study provides several new, converging lines of evidence for how hemispheric functional reorganization underlies adult foreign language learning. The right IFG, sensitive to acoustic features of speech sounds prepares successful learners to acquire novel speech categories at the earliest stage of learning. As learners' exposure to the foreign language increases, successful learners must generalize their knowledge about the novel speech categories across various speakers and syllables and ignore the within category acoustic differences. Excessive right-IFG involvement in speech processing at this stage might become detrimental. While our study does not address the causal influence between hemispheres, the relationship between the cross-hemispheric connectivity and right IFG disengagement indicates the possible inhibitory signal passed between hemispheres (Chiarello and Maxfield, 1996).

Our study sample had a median of 5 years of self-reported music training experience, which raises the question as to whether any of our findings could be biased due to participants' musical expertise. Prior research has shown long-term music training can benefit the processing of pitch patterns in native language (Magne et al., 2006) and learning of linguistic pitch contours of unfamiliar languages (Wong and Perrachione, 2007; Cooper and Wang, 2012; Smayda et al., 2015). No studies have yet investigated the impact of musical experience on holistic language learning outcomes. In the current study, the duration of musical training, a widely-used but arguably coarse measure of musical expertise, was not associated with Mandarin learning outcomes.

Another possibility is that musical expertise might potentially heighten participants' right hemisphere sensitivity for pitch processing and learning. Neuroimaging provides evidence that right frontotemporal pathways is crucial for pitch perception, production, and subsequent music training success (Loui et al., 2015; Herholz et al., 2016; Wollman et al., 2018; see Kraus and Chandrasekaran, 2010 for a review). In particular, right temporal cortex is specialized in fine-grained pitch-related computation (Hyde et al., 2008; Zatorre et al., 2012) and its structural enlargement is associated with musical proficiency (Gaser and Schlaug, 2003; Hyde et al., 2009). Given the tonal feature of Mandarin phonology, it is possible that the critical role of right IFG in Mandarin learning success in this study could also be attributed to the relatively high level of music expertise in our sample. However, we found a significant association between the functions of right IFG (both task activation and functional connectivity) and participants' foreign language learning aptitude measured by the MLAT, suggesting that the relationship between right IFG and Mandarin learning might be due to general language learning aptitude, rather than heightened sensitivity to pitch information in this sample. Nevertheless, our analyses do not rule out the possibility that music training experiences might have boosted the sensitivity of right IFG to foreign speech sounds in general.

Future studies are necessary to address a number of important questions, such as whether the right IFG is equally important for learning a non-tonal language and whether the right-to-left transition in function is specific to language training. Finally, the absence of a control group that did not learn Mandarin suggests caution in interpreting a causal relationship between training and the group-wide brain plasticity results shown in Fig. 3 (though, cf. Golestani and Zatorre, 2004; Myers and Swan, 2012, where analogous pre-to-post functional changes after training were observed), but does not affect our conclusions regarding the relationships between individual differences in brain measures and individual differences in learning outcomes from the Mandarin training intervention across individuals.

## 5. Conclusion

This study provides new converging evidence about the importance of the right hemisphere in speech perception for successful acquisition of multiple aspects of a real language in adults. Although the left hemisphere was increasingly recruited in most learners after training, individual differences in speech perception, immediate language learning attainment, and long-term retention of high-level foreign language skills were all related to differences in initial right-IFG engagement and subsequent right-IFG disengagement. The right-to-left transition was accompanied by an increase in resting-state functional connectivity between right IFG and left-hemisphere language regions. Taken together, these findings significantly refine the neurobiological model of ecological adult language learning and retention.

## Acknowledgment

We thank Amy S. Finn and Jennifer Minas for advice on the project design, and Valkyrie Felso for her contribution in data collection. We thank Atsushi Takahashi, Sheeba Arnold Anteraper, and Steven Shannon in Athinoula A. Martinos Imaging Center at McGovern Institute for Brain Research, Massachusetts Institute of Technology for their technical support. This work was supported by Defense Advanced Research Projects Agency DI-MISC 80508B to JDEG, Contract No. H98230-07-D-0175.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.03.008.

## References

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033–2044.
Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68, 255–278.
Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Linear Mixed-Effects Models Using "Eigen" and S4.
Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101.
Bitan, T., Lifshitz, A., Breznitz, Z., Booth, J.R., 2010. Bidirectional connectivity between hemispheres occurs at multiple levels in language processing but depends on sex. J. Neurosci. 30, 11576–11585.
Blumstein, S.E., Myers, E.B., Rissman, J., 2005. The perception of voice onset time: an fMRI investigation of phonetic category structure. J. Cogn. Neurosci. 17, 1353–1366.
Boersma, P., Weenink, D., 2001. Praat, a system for doing phonetics by computer. Glot Int. 5, 341–347.
Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. 1124, 1–38.
Burton, M., 2001. The role of inferior frontal cortex in phonological processing. Cogn. Sci. 25, 695–709.
Callan, D.E., Tajima, K., Callan, A.M., Kubo, R., Masaki, S., Akahane-Yamada, R., 2003. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. Neuroimage 19, 113–124.
Carroll, J., Sapon, S., 2002. Modern Language Aptitude Test MLAT : Manual 2002 Edition. Second Language Testing, Bethesda MD.
Chai, X.J., Berken, J.A., Barbeau, E.B., Soles, J., Callahan, M., Chen, J.-K., Klein, D., 2016. Intrinsic functional connectivity in the adult brain and success in second-language learning. J. Neurosci. 36, 755–761.
Chandrasekaran, B., Sampath, P.D., Wong, P.C.M., 2010. Individual variability in cue-weighting and lexical tone learning. J. Acoust. Soc. Am. 128, 456–465.
Chiarello, C., Maxfield, L., 1996. Varieties of interhemispheric inhibition, or how to keep a good hemisphere down. Brain Cogn. 30, 81–108.
Cook, N.D., 1984. Callosal inhibition: the key to the brain code. Behav. Sci. 29, 98–110.
Cooper, A., Wang, Y., 2012. The influence of linguistic and musical experience on Cantonese word learning. J. Acoust. Soc. Am. 131, 4756–4769.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179–194.
Díaz, B., Baus, C., Escera, C., Costa, A., Sebastián-Gallés, N., 2008. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. Proc. Natl. Acad. Sci. U. S. A. 105, 16083–16088.
Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A.C., Zilles, K., Amunts, K., 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage 36, 511–521.
Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure : why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. Unit. States Am. 113, 7900–7905.
Fedorenko, E., Behr, M.K., Kanwisher, N., 2011. Functional specificity for high-level linguistic processing in the human brain. Proc. Natl. Acad. Sci. Unit. States Am. 108, 16428–16433.
Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Gunther, M., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5, e15710.
Finn, A.S., Hudson Kam, C.L., 2008. The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. Cognition 108, 477–499.
Finn, A.S., Hudson Kam, C.L., Ettlinger, M., Vytlacil, J., D'Esposito, M., 2013. Learning language with the wrong neural scaffolding: the cost of neural commitment to sounds. Front. Syst. Neurosci. 7, 85.
Friederici, A.D., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91, 1357–1392.
Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. Neuroimage 23, 344–357.
Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., Li, X., 2002. A cross-linguistic FMRI study of spectral and temporal cues underlying phonological processing. J. Cogn. Neurosci. 14, 1076–1087.
Gaser, C., Schlaug, G., 2003. Brain structures differ between musicians and non-musicians. J. Neurosci. 23, 9240–9245.
Ghosh, S.S., Keshavan, A., Salvatore, J., Klein, A., 2012. BIPS: a framework for curating and executing brain imaging Pipelines. In: Front. Neuroinform. Conference Abstract: 5th INCF Congress of Neuroinformatics.
Golestani, N., Molko, N., Dehaene, S., LeBihan, D., Pallier, C., 2007. Brain structure predicts the learning of foreign speech sounds. Cerebr. Cortex 17, 575–582.
Golestani, N., Zatorre, R.J., 2004. Learning new sounds of speech: reallocation of neural substrates. Neuroimage 21, 494–506.
Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front. Neuroinf. 5, 13.
Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72.
Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.
Herholz, S.C., Coffey, E.B.J., Pantev, C., Zatorre, R.J., 2016. Dissociation of neural networks for predisposition and for training-related plasticity in auditory-motor learning. Cerebr. Cortex 26, 3125–3134.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.
Hinkley, L.B.N., Marco, E.J., Brown, E.G., Bukshpun, P., Gold, J., Hill, S., Findlay, A.M., Jeremy, R.J., Wakahiro, M.L., Barkovich, A.J., Mukherjee, P., Sherr, E.H., Nagarajan, S.S., 2016. The contribution of the corpus callosum to language lateralization. J. Neurosci. 36, 4522–4533.
Hoeft, F., McCandliss, B.D., Black, J.M., Gantman, A., Zakerani, N., Hulme, C., Lyytinen, H., Whitfield-Gabrieli, S., Glover, G.H., Reiss, A.L., Gabrieli, J.D.E., 2011. Neural systems predicting long-term outcome in dyslexia. Proc. Natl. Acad. Sci. U. S. A. 108, 361–366.
Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biom. J. 50, 346–363.
Hsieh, L., Gandour, J., Wong, D., Hutchins, G.D., 2001. Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang. 76, 227–252.
Hyde, K.L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A.C., Schlaug, G., 2009. Musical training shapes structural brain development. J. Neurosci. 29, 3019–3025.

Hyde, K.L., Peretz, I., Zatorre, R.J., 2008. Evidence for the role of the right auditory cortex in fine pitch resolution. Neuropsychologia 46, 632–639.
Kaufman, A.S., Kaufman, N.L., 2004. Kaufman Brief Intelligence Test, second ed. Pearson.
Kormos, J., Sáfár, A., 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. Biling. Lang. Cognit. 11, 261–271.
Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. Nat. Rev. Neurosci. 11, 599–605.
Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 979–1000.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.
Linck, J. a., Hughes, M.M., Campbell, S.G., Silbert, N.H., Tare, M., Jackson, S.R., Smith, B.K., Bunting, M.F., Doughty, C.J., 2013. Hi-lab: a new measure of aptitude for high-level language proficiency. Lang. Learn. 63, 530–566.
Liu, X., 2010. New Practical Chinese Reader, second ed., vol. 1. Beijing Language Culture University Press.
Loui, P., Demorest, S.M., Pfordresher, P.Q., Iyer, J., 2015. Neurological and developmental approaches to poor pitch perception and production. Ann. N. Y. Acad. Sci. 1337, 263–271.
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. U. S. A. 103, 19558–19563.
Magne, C., Schön, D., Besson, M., 2006. Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. J. Cogn. Neurosci. 18, 199–211.
Mårtensson, J., Eriksson, J., Bodammer, N.C., Lindgren, M., Johansson, M., Nyberg, L., Lövdén, M., 2012. Growth of language-related brain areas after foreign language learning. Neuroimage 63, 240–244.
Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63, 1144–1153.
Morey, R.D., 2008. Confidence intervals from normalized data: a correction to Cousineau. Tutor. Quant. Methods Psychol. 4, 61–64.
Moulines, E., Charpentier, F., 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun. 9, 453–467.
Mueller, J.L., Friederici, A.D., Männel, C., 2012. Auditory perception at the root of language learning. Proc. Natl. Acad. Sci. U. S. A. 109, 15953–15958.
Murtagh, L., van der Slik, F., 2004. Retention of Irish skills: a longitudinal study of a school-acquired second language. Int. J. Biling. 8, 279–302.
Myers, E.B., 2014. Emergence of category-level sensitivities in non-native speech sound learning. Front. Neurosci. 8, 1–11.
Myers, E.B., Blumstein, S.E., Walsh, E., Eliassen, J., 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychol. Sci. 20, 895–903.
Myers, E.B., Swan, K., 2012. Effects of category learning on neural sensitivity to non-native phonetic categories. J. Cogn. Neurosci. 24, 1695–1708.
Naeser, M.A., Martin, P.I., Ho, M., Treglia, E., Kaplan, E., Bashir, S., Pascual-Leone, A., 2012. Transcranial magnetic stimulation and aphasia rehabilitation. Arch. Phys. Med. Rehabil. 93, 26–34.
Naeser, M.A., Martin, P.I., Theoret, H., Kobayashi, M., Fregni, F., Nicholas, M., Tormos, J.M., Steven, M.S., Baker, E.H., Pascual-Leone, A., 2011. TMS suppression of right pars triangularis, but not pars opercularis, improves naming in aphasia. Brain Lang. 119, 206–213.
Nicolo, P., Fargier, R., Laganaro, M., Guggisberg, A.G., 2016. Neurobiological correlates of inhibition of the right broca homolog during new-word learning. Front. Hum. Neurosci. 10, 371.
Nir, Y., Mukamel, R., Dinstein, I., Privman, E., Harel, M., Fisch, L., Gelbard-Sagiv, H., Kipervasser, S., Andelman, F., Neufeld, M.Y., Kramer, U., Arieli, A., Fried, I., Malach, R., 2008. Interhemispheric correlations of slow spontaneous neuronal fluctuations revealed in human sensory cortex. Nat. Neurosci. 11, 1100–1108.
Norman-Haignere, S., Kanwisher, N., McDermott, J.H., 2013. Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. J. Neurosci. 33, 19451–19469.
Perrachione, T.K., Ghosh, S.S., 2013. Optimized design and analysis of sparse-sampling FMRI experiments. Front. Neurosci. 7, 55.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154.
Qi, Z., Han, M., Garel, K., Chen, E.S., Gabrieli, J.D.E., 2015. White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. J. Neurolinguistics 33, 14–28.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98, 676–682.
Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., Kuhl, P.K., 2005. Neural patterns to speech and vocabulary growth in American infants. Neuroreport 16, 495–498.
Robinson, P., 2001. Individual differences, cognitive abilities, aptitude complexes, and learning conditions in second language acquisition. Sec. Lang. Res. 17, 368–392.
Schlegel, A. aA., Rudelson, J.J., Tse, P.U., 2012. White matter structure changes as adults learn a second language. J. Cogn. Neurosci. 24, 1664–1670.
Sebastián-Gallés, N., Díaz, B., 2012. First and second language speech perception: graded learning. Lang. Learn. 62, 131–147.
Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn. Reson. Med. 67, 1210–1224.
Siegel, J.S., Power, J.D., Dubis, J.W., Vogel, A.C., Church, J.A., Schlaggar, B.L., Petersen, S.E., 2014. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Hum. Brain Mapp. 35, 1981–1996.
Silbert, N.H., Smith, B.K., Jackson, S.R., Campbell, S.G., Hughes, M.M., Tare, M., 2015. Non-native phonemic discrimination, phonological short term memory, and word learning. J. Phonet. 50, 99–119.
Silver, M., Montana, G., Nichols, T.E., 2011. Alzheimer's Disease Neuroimaging Initiative the ADN, False positives in neuroimaging genetics using voxel-based morphometry data. Neuroimage 54, 992–1000.
Smayda, K.E., Chandrasekaran, B., Maddox, W.T., 2015. Enhanced cognitive and perceptual processing: a computational basis for the musician advantage in speech learning. Front. Psychol. 6, 682.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (Suppl. 1), S208–S219.
Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann, C.F., Jenkinson, M., Andersson, J., Glasser, M.F., Van Essen, D.C., Feinberg, D. aA., Yacoub, E.S., Ugurbil, K., 2012. Temporally-independent functional modes of spontaneous brain activity. Proc. Natl. Acad. Sci. U. S. A. 109, 3131–3136.
Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., Brandeis, D., Dierks, T., 2012. Structural plasticity in the language system related to increased second language proficiency. Cortex 48, 458–465.
Thesen, S., Heid, O., Mueller, E., Schad, L.R., 2000. Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. Magn. Reson. Med. 44, 457–465.
van der Kouwe, A.J.W., Benner, T., Salat, D.H., Fischl, B., 2008. Brain morphometry with multiecho MPRAGE. Neuroimage 40, 559–569.
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, J.A., Jongman, A., Hirsch, J., 2003. fMRI evidence for cortical modification during learning of Mandarin lexical tone. J. Cogn. Neurosci. 15, 1019–1027.
Whitfield-Gabrieli, S., Nieto-Castanon, A., 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect. 2, 125–141.
Wollman, I., Penhune, V., Segado, M., Carpentier, T., Zatorre, R.J., 2018. Neural network retuning and neural predictors of learning success associated with cello training. Proc. Natl. Acad. Sci. U. S. A. 115, E6056–E6064.
Wong, F.C.K., Chandrasekaran, B., Garibaldi, K., Wong, P.C.M., 2011. White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. J. Neurosci. 31, 8780–8785.
Wong, P.C.M., Parsons, L.M., Martinez, M., Diehl, R.L., 2004. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. J. Neurosci. 24, 9153–9160.
Wong, P.C.M., Perrachione, T.K., 2007. Learning pitch patterns in lexical identification by native English-speaking adults. Appl. Psycholinguist. 28, 565–585.
Wong, P.C.M., Perrachione, T.K., Parrish, T.B., 2007. Neural characteristics of successful and less successful speech and word learning in adults. Hum. Brain Mapp. 28, 995–1006.
Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. Neuroimage 45, S173–S186.
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.
Yang, J., Gates, K.M., Molenaar, P., Li, P., 2014. Neural changes underlying successful second language word learning: an fMRI study. J. Neurolinguistics 1–21.
Zatorre, R.J., Delhommeau, K., Zarate, J.M., 2012. Modulation of auditory cortex response to pitch variation following training with microtonal melodies. Front. Psychol. 3, 544.
Zatorre, R.J., Gandour, J.T., 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 1087–1104.