

# Human inferior colliculus activity relates to individual differences in spoken language learning

Bharath Chandrasekaran,<sup>1</sup> Nina Kraus,<sup>2</sup> and Patrick C. M. Wong<sup>3</sup>

<sup>1</sup>*Department of Communication Sciences and Disorders, Center for Perceptual Systems, Institute for Neuroscience, The University of Texas at Austin, Austin, Texas;* <sup>2</sup>*The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Department of Neurobiology and Physiology, Department of Otolaryngology—Head and Neck Surgery, and The Hugh Knowles Hearing Center, Northwestern University, Evanston;* and <sup>3</sup>*The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Department of Otolaryngology—Head and Neck Surgery, and The Hugh Knowles Hearing Center, Northwestern University, Evanston, Illinois*

Submitted 12 October 2011; accepted in final form 24 November 2011

**Chandrasekaran B, Kraus N, Wong PC.** Human inferior colliculus activity relates to individual differences in spoken language learning. *J Neurophysiol* 107: 1325–1336, 2012. First published November 30, 2011; doi:10.1152/jn.00923.2011.—A challenge to learning words of a foreign language is encoding nonnative phonemes, a process typically attributed to cortical circuitry. Using multimodal imaging methods [functional magnetic resonance imaging-adaptation (fMRI-A) and auditory brain stem responses (ABR)], we examined the extent to which pretraining pitch encoding in the inferior colliculus (IC), a primary midbrain structure, related to individual variability in learning to successfully use nonnative pitch patterns to distinguish words in American English-speaking adults. fMRI-A indexed the efficiency of pitch representation localized to the IC, whereas ABR quantified midbrain pitch-related activity with millisecond precision. In line with neural “sharpening” models, we found that efficient IC pitch pattern representation (indexed by fMRI) related to superior neural representation of pitch patterns (indexed by ABR), and consequently more successful word learning following sound-to-meaning training. Our results establish a critical role for the IC in speech-sound representation, consistent with the established role for the IC in the representation of communication signals in other animal models.

pitch; functional magnetic resonance imaging-adaptation; repetition suppression; inferior colliculus

THE INFERIOR COLLICULUS (IC), the primary auditory midbrain structure, has topographical maps of frequency and temporal rate information (Baumann et al. 2011; Schreiner and Langner 1988, 1997). These topographical maps suggest a critical role for the IC in the representation of spectrotemporal variations inherent in communication signals. Given that the IC is a neural convergence zone (it receives ascending inputs from lower auditory nuclei and top-down corticofugal connections from the auditory cortex and limbic regions), the IC has been hypothesized to play a critical role in auditory learning (Bajo et al. 2010). Animal models show that short-term auditory training modulates activity in the IC (Bajo et al. 2010; Suga 2008; Suga and Ma 2003; Sun et al. 2007). Furthermore, impairing IC function using lesion methods severely disrupts auditory learning (Bajo et al. 2010). In humans, neuroimaging and associated studies examining auditory and language learning have exclusively focused on cortical structures (Jäncke et al.

2001; Peach and Wong 2004; Perrachione and Wong 2007; Wang et al. 2003; Wong et al. 2007a, 2011). Although individual variability in neurophysiological responses from the auditory cortex is predictive of auditory learning success (Wong et al. 2007a), the extent to which IC neural signals have a similar predictive value has not been established.

Animal models demonstrate the existence of neurons that are highly selective to the direction of frequency modulations (FMs) in the IC (Williams and Fuzessery 2010; Zhang et al. 2003). The direction of FMs is a critical parameter in human (and animal) communicative signals (Zhang et al. 2003). In certain languages of the world called “tone languages,” FM sweeps within a syllable can change word meaning. For example, in Mandarin Chinese, a prominent tone language, the syllable /ma/ spoken with level (T1), rising (T2), dipping (T3), or falling (T4) pitch pattern can mean “mother,” “hemp,” “horse,” and “to scold,” respectively. Previous studies have found considerable variability in the ability to learn to use pitch patterns to contrast words in nonnative speakers (Chandrasekaran et al. 2007, 2010; Wang et al. 2003; Wong et al. 2007a). It has been hypothesized that such learning differences may originate from a differential ability to encode pitch direction, an important cue used by native speakers to disambiguate pitch patterns (Chandrasekaran et al. 2010; Wang et al. 2003). Since the IC contains FM direction-sensitive neurons that are important for processing tonal sweeps (Andoni et al. 2007; Brimijoin and O’Neill 2005) and is highly sensitive to the direction of FM sweeps (Andoni et al. 2007; Brimijoin and O’Neill 2005), we predicted that individual differences in the pretraining IC encoding of these complex time-varying patterns might relate to success in mapping these patterns onto words.

We used a functional magnetic resonance imaging-adaptation (fMRI-A) paradigm to probe pitch pattern-specific activity from the IC. In fMRI-A, a stimulus feature (e.g., pitch) is repeated, which typically results in a reduction in blood oxygen level-dependent (BOLD) activity in regions sensitive to the stimulus feature that is repeated. This reduction is argued to reflect increased neural processing efficiency of the stimulus (Grill-Spector et al. 2006). The “repeat” condition is then compared with a “change” condition, during which the stimulus feature is continuously varied. The difference between the two conditions is used as an index of stimulus-specific neural activity. Several models have been proposed to explain the reduction in BOLD response (repetition suppression, RS) with

Address for reprint requests and other correspondence: P. C. Wong, Northwestern Univ., 2240 Campus Drive North, Evanston, IL 60208 (e-mail: pwong@northwestern.edu).

prior experience. Broadly, models view RS either as an automatic process occurring locally (due to fatigue or sharpening of stimulus representation with repetition) or as a top-down process driven by perceptual expectations (Summerfield et al. 2008). Although current models differ on the reasons for RS (change > repeat), they converge on the fact that suppression is associated with improved neural encoding of the signal with repetition (Grill-Spector et al. 2006). In contrast, neither model makes predictions as to why repetition sometimes causes an increase in fMRI signal (repetition enhancement, RE). For example, whereas word repetition has been shown to reduce neural activity in the brain regions sensitive to words (RS), listening to repeating pseudowords increases the fMRI signal in these regions (RE) (Gagnepain et al. 2008). RE is often elicited by novel stimuli but is not always driven by novelty (Henson et al. 2000); a recent study showed that repetition of degraded familiar stimuli resulted in RE, whereas nondegraded familiar stimuli resulted in RS (Turk-Browne et al. 2007). This suggests that the direction of repetition effects (RS vs. RE) may be determined by the quality of neural activity elicited by the sensory input. Although a number of factors are known to modulate the extent of neural repetition effects (Henson 2003; Horner and Henson 2008), to our knowledge, no study has examined the behavioral ramification of individual differences in the direction (RS vs. RE) of neural repetition effects. Here we examine the hypothesis that individual differences in the direction of neural repetition effects reflect differences in the quality of stimulus encoding in neural ensembles within the IC. In line with “sharpening” models, we predict that individuals who show RS will have enhanced (more fine-tuned) representation of relevant stimulus parameters (pitch direction). Consequently, we propose that these representational differences will yield differential success in auditory learning.

To test these hypotheses, we elicited IC BOLD responses to nonnative pitch patterns presented in three conditions. In the repeat condition, the same tone was repeated four times (e.g., T1, T1, T1, T1). Tone-change conditions had four different tones (e.g., T4, T1, T2, T3), and a control condition (speaker change) used the same tones but were recorded from multiple talkers (e.g., T2s1, T2s2, T2s3, T2s4). Participants were then divided into two groups, individuals who showed RS (tone change > repeat) and those who showed RE (repeat > tone change). To examine representational differences between the two groups, we compared their brain stem frequency-following responses (FFRs; recorded in a separate session) with a rising pitch pattern. The FFRs are a scalp-recorded component of the auditory brain stem response (ABR) that reflects ensemble brain stem phase-locked activity to the eliciting stimulus (presented repetitively) with millisecond precision (Chandrasekaran and Kraus 2010; Skoe and Kraus 2010). Animal studies suggest IC as being the source of the scalp-recorded FFR (Chandrasekaran and Kraus 2010; Galbraith et al. 2001; Smith et al. 1975, 1978). Unlike cortical responses, FFRs allow the near-exact quantification between the stimulus properties and the electrophysiological response. The close similarity between the stimulus and the FFRs allow the examination of training-induced changes to specific stimulus properties (Song et al. 2011). In the current study, FFRs allow the examination of representational differences between the RE and RS groups, as well as training-induced modifications in midbrain pitch representation. FFRs were recorded to a 278-ms rising (T2) pitch

pattern (Fig. 1), which had a flat slope in the first half of the stimulus (steady state: fundamental frequency rose from 110 to 119 Hz) and a rising slope in the second half of the stimulus (dynamic: fundamental frequency rose from 119 to 163 Hz). Consistent with previous studies, we separately analyzed FFRs to the equal-duration steady-state and dynamic portions of the stimuli (Krishnan et al. 2009; Strait et al. 2009). This allowed us to quantify group differences in steady-state vs. dynamic pitch encoding in the IC. Second, we examined the extent to which the direction of fMRI repetition effects (IC RS vs. IC RE) predicted success in learning to use these pitch patterns after nine sessions of auditory training. Finally, we elicited FFRs posttraining to examine the extent to which auditory training modulated IC dynamic pitch representation in the two groups.

## MATERIALS AND METHODS

**Participants.** Twenty adult (mean age 25.9 yr, SD 4.6 yr), right-handed, native speakers of American English took part in this study. Participants were all monolingual English speakers and reported that English was the only language that they spoke fluently and used for communication. Participants reported no prior exposure to any tone language. Two of these participants did not complete their training regimen, so analyses were restricted to 18 participants. Participants had no significant musical expertise (<6 yr of continuous musical training, starting after the age of 7 yr). All passed a hearing-screening test (thresholds of <25 dB HL at frequencies of 500, 1,000, 2,000, and 4,000 Hz). Our protocol was approved by the Institutional Review Board (IRB) of Northwestern University. Participants provided written informed consent in accordance with guidelines set by the Northwestern University IRB. Participants also underwent a battery of cognitive tests that included an assessment of auditory working memory, nonverbal IQ, and phonological processing ability. Participants underwent two pretraining fMRI scans (localizer, adaptation) and an electroencephalogram (EEG) recording prior to nine sessions of sound-to-meaning auditory training. Posttraining participants underwent a second EEG scan.

**Stimuli.** Stimuli used in the functional localizer task, fMRI-A experiment, and sound-to-meaning training paradigm were identical to stimuli used in a prior experiment examining cue-weighting differences in good and poor learners (Chandrasekaran et al. 2010). For the functional localizer task, two male and two female speakers of Mandarin Chinese each produced five Mandarin vowels /a/, /i/, /o/, /e/, and /y/ with Mandarin tone 1 (level tone). After this, hybrid stimuli were created using the pitch-synchronous overlap and add (PSOLA) algorithm implemented using Praat software. For each vowel and talker, three sets of stimuli were synthesized with level (T1), rising (T2), and falling (T4) pitch patterns. T3, which contains both falling and rising (phonetically described as “low dipping”) was not used because this pitch pattern was hard to label by nonnative speakers. Stimuli/talkers used in this task were not used in the fMRI-A task or in the sound-to-meaning training paradigm. In total, 60 stimuli were created (5 vowels  $\times$  3 tones  $\times$  4 talkers).

For the fMRI-A experiment and sound-to-meaning training paradigm, stimuli consisted of native pitch patterns produced by Mandarin speakers (2 male, 2 female) superimposed onto 6 English pseudowords produced by native American English speakers (2 male, 2 female), resulting in 96 hybrid stimuli that contained native phonotactics and foreign pitch patterns (6 pseudowords  $\times$  4 tones  $\times$  4 talkers). To create the hybrid stimuli, five native Mandarin (3 male, 2 female) speakers each produced the syllable /mi/ with the four different Mandarin tonal patterns (/mi-T1/, /mi-T2/, /mi-T3/, /mi-T4/). Four native speakers of American English produced six syllables with English pseudowords. For creating the hybrid stimuli, the duration of the pitch contour in the /mi/ syllable and the pseudoword were

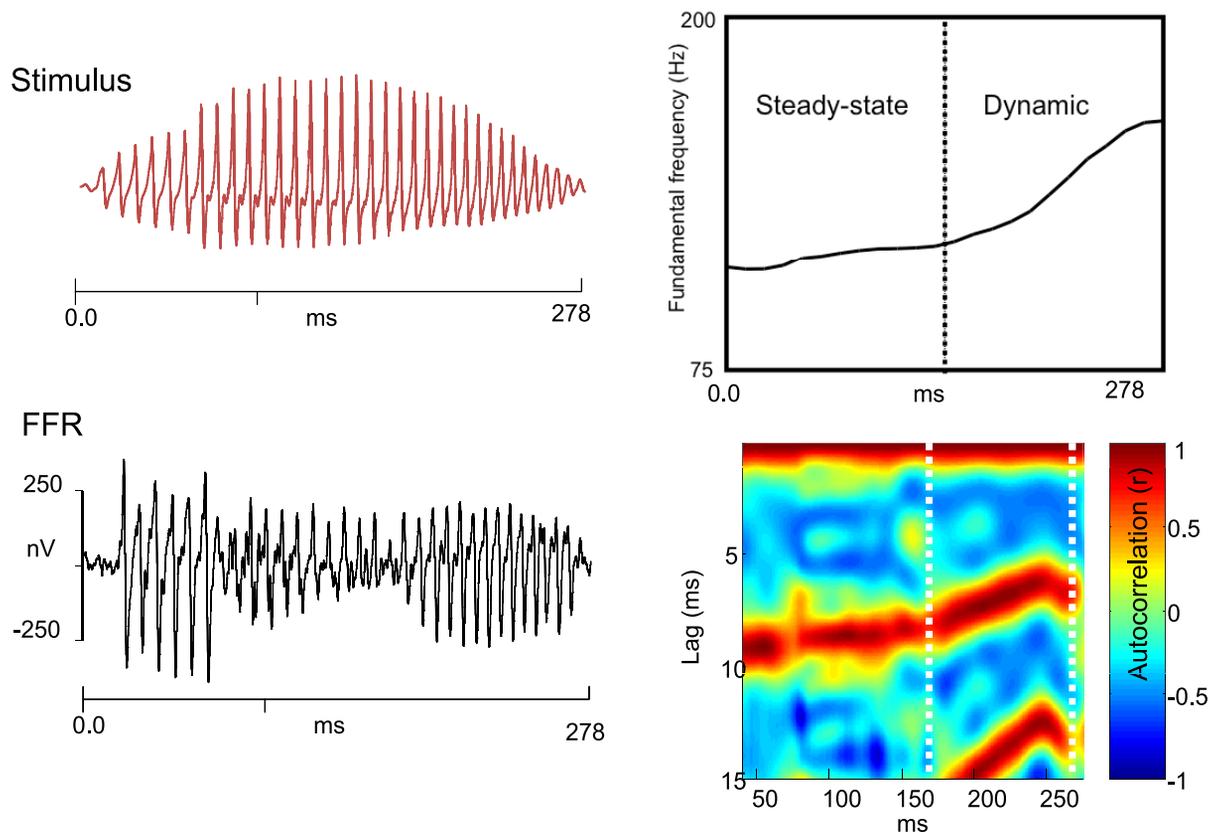


Fig. 1. *Top*: waveform (*left*) and fundamental frequency ( $F_0$ ; *right*) of the stimulus /mi/ used to elicit frequency-following responses (FFR). The stimulus pitch is largely invariant during the first half (steady state) and dynamic (rising) during the second half of the stimulus. *Bottom*: The grand-average FFR (*left*) closely mimics the stimulus with millisecond precision. Autocorrelation of the response (*right*) shows phase-locked activity to the stimulus pitch contour. Color indicates the magnitude of the autocorrelation, with warmer colors signifying higher magnitudes than cooler colors.

calculated. The duration of the voiced portion of the consonant-vowel (CV) syllable was then matched in duration to the voiced portion of the consonant-vowel-consonant (CVC) syllable using Praat software. Pitch patterns extracted from Mandarin productions were sex-matched with the native English pseudowords; that is, male and female pitch patterns were superimposed on male and female pseudoword production, respectively. The PSOLA method implemented in the Praat software was used to replace the pitch contour of /mi/ onto the CVC syllable. The 24 tokens per talker were created by superimposing Mandarin tones on the English base pseudowords, yielding a total of 96 tokens used in the training study. All four Mandarin pitch patterns were used (T1, phonetically described as “high-level,” T2 as “low-rising,” T3 as “low-dipping,” and T4 as “high-falling”). For the fMRI-A experiment, stimuli were presented in sets of four across three different conditions. In all conditions, the same pseudoword (e.g., “Pesh”) was repeated. During repeat trials, the same pitch pattern was repeated (e.g., T2, T2, T2, T2). During tone-change trials, different pitch patterns were presented (e.g., T2, T1, T3, T4). A third condition (speaker change) was introduced as a control. In this condition different speakers (2 male, 2 female) produced the same pitch pattern (e.g., T2, T2, T2, T2). Thus the global pitch pattern remained the same across all four talkers, but individual voice pitch was different for each speaker. Speaker change acts as an ideal control because although it involves pitch change, the change is more in terms of pitch height than pitch direction [i.e., a rising tone produced by a male talker and female talker is perceived as “rising,” although the average pitch (pitch height) is substantially lower in the male talker]. Critical features for tone processing (pitch direction) vary rapidly within the syllable in the tone-change condition. In contrast, in the speaker-change condition, pitch is more stable over time (although

speakers themselves have different average pitch). Stimuli used in the fMRI-A task were the same as those used in the sound-to-meaning training program.

The stimulus used in the ABR experiment was identical to the one used in a previous experiment that compared musicians’ and nonmusicians’ brain stem encoding of pitch patterns (Wong et al. 2007b). This stimulus consisted of the syllable [mi] with T2 (rising) pitch pattern, produced by a male speaker of Mandarin Chinese, and was 278 ms in duration. This stimulus was used for a number of reasons. During the first half of the stimulus, the pitch pattern is relatively constant (steady state, Fig. 1). During the second half of the stimulus, pitch dramatically rises (dynamic, Fig. 1). Previous cross-language (Chinese, English) studies examining ABRs to T2 found that maximum cross-language differences between native and nonnative speakers occur in the rising portion of T2. Since our hypothesis is that difficulty encoding pitch direction (and not pitch per se) is related to individual variability in sound-to-meaning learning, we used this pitch pattern to elicit the ABR. By comparing the relative encoding difference between the steady-state portion and the dynamic portion of the neural response to the stimulus, we were able to address the specificity of encoding differences between groups. This token /mi-T2/ was not a part of the training stimuli and allowed the examination of the effects of training on brain stem encoding (in contrast to exposure-driven modifications in encoding).

*fMRI IC localizer task.* We used a functional localizer task (pitch direction identification) to create an IC region of interest (ROI). This independently derived IC ROI was used to determine IC repetition effects in the fMRI-A experiment. The pitch direction identification task was used to localize the IC because animal studies show that the IC is strongly activated by FM sweeps. BOLD responses were

collected using a sparse sampling design (Belin et al. 2000; Chandrasekaran et al. 2011; Hall et al. 1999). Participants heard one syllable (vowels with superimposed linear rising, level, or falling pitch patterns synthesized from 4 different talkers who were not used in the training study) at a time and were visually presented three different pictures (i.e.,  $\rightarrow$  = level,  $\uparrow$  = rising, and  $\downarrow$  = falling) depicting the pitch pattern of the auditory stimulus. Participants performed a pitch direction judgment task, deciding whether pitch was rising, level, or falling on each trial (40 trials in total). Silent trials ( $n = 40$ ), during which no stimuli were presented, were interspersed randomly with task trials. Participants were instructed to indicate their choice on a MRI-compatible button box. Subjects were imaged at the first 2 s of a TR (repetition time) followed by 12 s (6 stimuli) of stimulus presentations when no scanning took place. For all participants, the fMRI IC localizer task was performed before the fMRI-A experiment.

**Definition of IC ROI.** IC ROIs for the fMRI-A experiment were established with the following steps. In *step 1*, from the localizer task, voxels significantly active for the task > silence contrast at the second level were extracted using an uncorrected threshold of  $P = 0.001$  (voxel extent >40 voxels), consistent with IC functional mapping methods used in previous studies examining subcortical auditory activity (Griffiths et al. 2001; von Kriegstein et al. 2008). The left IC (and not the right IC) was significantly active in this step (Fig. 2). In *step 2*, we then created a left IC mask [ $5\text{-mm}^3$  sphere centered on the peak IC activation reported in a previous study that examined subcortical encoding of pitch periodicity (Griffiths et al. 2001)]. The final subcortical ROI was conservatively calculated as clusters significant in *step 1* enclosed within the  $5\text{-mm}^3$  masks generated in *step 2*. For

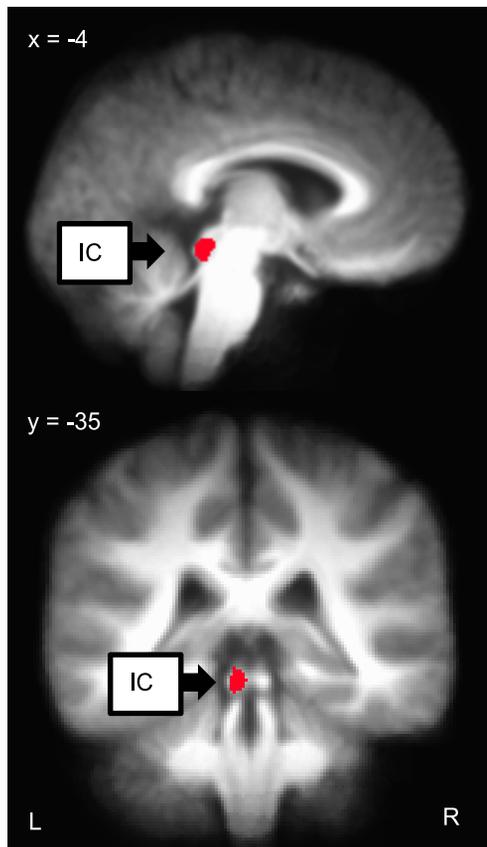


Fig. 2. The left inferior colliculus (IC) was significantly activated by the auditory localizer task ( $P < 0.001$ , uncorrected, voxel extent >40). Significant voxels (spatial extent = 246 voxels) located within IC regions of interest (ROI) from a previous study (Griffiths et al. 2001) were used to create the IC mask (in red).

this ROI, mean %signal change was calculated for each contrast of interest for every participant. These were then converted to  $z$  scores.

**fMRI-A design.** Participants performed an orthogonal loudness judgment task in the scanner while listening to speech sounds presented in two different conditions. In the repeat condition, the speech sound containing a tone was repeated four times (e.g., T1, T1, T1, T1) within the TR of 12 s. In tone-change conditions, the tone varied within the TR of 10 s (e.g., T4, T1, T2, T3). In the speaker-change condition, the speech sounds contained the same tone but multiple talkers. Irrespective of condition, the task was to indicate if the second stimulus was louder or softer than the first stimulus. The first and third stimuli of the four-stimulus sequence were normalized to a fixed amplitude level. The second and fourth stimuli were either 5 dB louder or 5 dB softer. This orthogonal task was used to ensure attentiveness throughout the experiment. Loudness judgment tasks do not interfere with pitch processing, as shown in previous studies (Chandrasekaran et al. 2011; Church and Schacter 1994). Since participants performed at ceiling (>90% accuracy) during the loudness judgment task for all three conditions, task performance was not considered as a factor during analyses. The sparse sampling design allows participants to listen to the stimuli without the interference of scanner noise and provides adequate time for the scanner noise-induced hemodynamic response (HR) to reduce and not overlap with HR for the auditory stimuli (Belin et al. 2000; Chandrasekaran et al. 2011; Hall et al. 1999) (Fig. 3). There were 40 blocks per condition and 40 blocks of silent trials. The null trials were used to establish baseline.

**Participant grouping.** Participants were then divided into two groups based on the tone change > repeat contrast (IC RS vs. IC RE). Individuals who had positive activation in the tone change > repeat contrast were grouped into the IC suppression group, and those who had negative activation in the tone change > repeat condition were classified as showing IC enhancement (Fig. 3). As a control, a second grouping was created by dividing participants into two groups based on the speaker change > repeat contrast (IC speaker-RS vs. IC speaker-RE). Repeated-measures ANOVA (rmANOVA) were conducted to examine the extent to which the two groups differed on the basis of fidelity and auditory learning success.

**MRI acquisition and data analysis.** MR images were acquired using a Siemens 3T Trio MRI scanner. Apart from the difference in TR (14 s for the fMRI localizer task and 12 s for the fMRI-A experiment), the procedures were identical for both the localizer and fMRI-A paradigms. For each participant, a high-resolution, T1-weighted three-dimensional volume was acquired axially [MP-RAGE; TR = 2,300 ms; echo time (TE) = 3.36 ms, flip angle =  $9^\circ$ , inversion time (TI) = 900 ms, matrix size =  $256 \times 256$ , field of view (FOV) = 22 cm, slice thickness = 1 mm]. For the functional localizer task, T2\*-weighted images acquired axially with an echo-planar imaging (EPI) pulse sequence [TE = 30 ms, TR = 14 s, flip angle =  $90^\circ$ , in-plane resolution =  $3.4375 \times 3.4375$  mm, 38 slices with slice thickness = 3 mm (without gap between slices)] were acquired in an interleaved measurement. For the fMRI-A experiment, T2\*-weighted images were acquired axially using a susceptibility weighted EPI pulse sequence [TE = 30 ms, TR = 12 s, flip angle =  $90^\circ$ , in-plane resolution =  $3.4375 \times 3.4375$  mm, 38 slices with slice thickness = 3 mm (without gap between slices)] were acquired in an interleaved measurement. The functional MR images (for both the functional localizer and fMRI-A) were analyzed using AFNI software (Cox 1996). Images were corrected for motion and slice time, and spatial smoothing [full-width at half-maximum (FWHM) = 6 mm] was performed, followed by linear detrending and resampling to a resolution of  $3\text{ mm}^3$ . Square waves modeling the events of interest were created as extrinsic model waveforms of the task-related HR. The waveforms of the modeled events were then used as regressors in a multiple linear regression of the voxel-based time series. Normalized beta values signifying the fit of the regressors to the functional scanning series, voxel-by-voxel for each condition, were used for

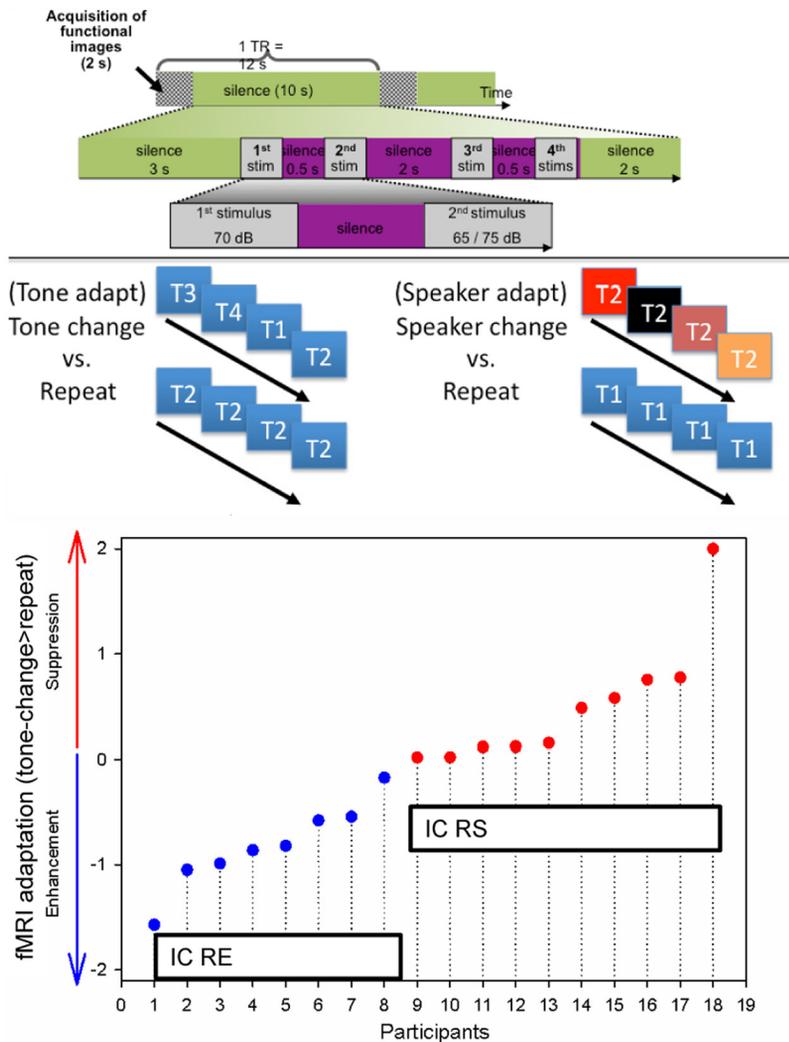


Fig. 3. Functional magnetic resonance imaging (fMRI) design. Participants listened to 4 sounds presented repetitively across 3 conditions as blood oxygen level-dependent (BOLD) responses were collected using a sparse sampling design. Irrespective of conditions, participants judged whether the second sound was louder or softer than the first sound. In the first condition (repeat), participants listened to the same speaker repeat the same pseudoword (e.g., /Pesh/) with the same pitch pattern (e.g., T2, T2, T2, T2). In a second condition (tone change), participants listened to the same speaker repeat the same pseudoword with different pitch patterns (e.g., T3, T4, T1, T2). In the third (control) condition (speaker change), participants listened to different speakers (s1, s2, s3, s4) repeat the same pseudoword with the same pitch pattern (e.g., T2s2, T2s4, T2s3, T2s1). Graph shows averaged %signal change, extracted from the IC ROI for the contrast tone change > repeat and converted to  $z$  scores ( $y$ -axis). Considerable variability was seen in the extent of tone adaptation. A number of participants showed positive activity in the tone change > repeat contrast; others showed greater activity in the reverse contrast. These participants were divided into groups based on those who showed tone-related repetition suppression indexed by the IC fMRI signal (IC RS; in red) and those who showed tone-related repetition enhancement (RE) indexed by the IC fMRI signal (IC RE; in blue), respectively.

group analyses. Anatomical and functional images from each subject were normalized to a standard stereotaxic template (ICBM 152).

**ABR recording.** During the ABR recordings, subjects watched a movie of choice (volume muted, English subtitles provided) while the syllable /mi/ with a rising (T2) pitch pattern produced by a native speaker of Mandarin (Fig. 1) was presented binaurally via insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL) at 75 dB SPL. Stimuli were presented with a variable interstimulus interval between 71.50 and 104.84 ms. Responses were recorded using Scan 4.3 (Neuroscan; Compumedics) with Ag-AgCl scalp electrodes, recorded from Cz (active) to linked earlobes (reference), with the forehead as ground. Contact impedance was <5 k $\Omega$  for all electrodes. Artifact-free sweeps (1,500) were collected from each polarity (total 3,000 sweeps) with a sampling rate of 20 kHz. Responses were bandpass filtered from 80 to 1,000 Hz, 12 dB/octave, and trials with activity greater than  $\pm 35 \mu\text{V}$  were rejected. Waveforms were averaged 45 ms before onset and 16.5 ms after offset of the stimulus, consistent with previous studies (Song et al. 2008; Wong et al. 2007b). Responses of alternating polarity were then added together to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Chandrasekaran and Kraus 2010).

**ABR analyses.** For each participant, we calculated two FFR measures: root mean square amplitude of the response (FFR amplitude, measured from 10 to 288 ms) and stimulus-to-response correlation. The stimulus-to-response correlation measure, which provides a global measure of pitch-tracking accuracy, represents the extent to which the stimulus and response pitch contours are correlated (Pear-

son's correlation coefficient  $r$  between the stimulus and response F0 contours) such that a value of 1 would indicate that the contours were identical (closer to 1 indicates more faithful pitch encoding). This measure represents both the strength and direction of the linear relationship between the two signals. We also examined signal-to-noise ratio to ensure that the quality of the recording was consistent across the two groups of participants (no significant differences were found,  $P > 0.6$ , paired sample  $t$ -test). To obtain the FFR pitch contour, we used a sliding-window analysis procedure in which 40-ms bins of the FFR were analyzed in the lag (autocorrelation) domain. The 40-ms sliding window was shifted in 1-ms steps across the two segments of the stimulus (Fig. 1, steady-state portion from 0 to 139 ms and time-varying portion from 139 to 278 ms). To estimate the period of the quasiperiodic waveform over time, each bin was cross-correlated with itself (autocorrelated). The maximum autocorrelation was found over a lag range of 5–10 ms, a range that encompasses the F0 contour of T2. The maximum (peak) autocorrelation value (expressed as a value between  $-1$  and  $1$ ) was recorded for each bin, with higher values indicating more periodic time frames (Fig. 1). To generate the pitch contour, the lag (i.e., associated with each maximum autocorrelation) was converted to frequency (lag and frequency are reciprocally related). The same sliding-window procedure was applied to the stimulus waveforms to derive the respective stimulus pitch contour. All pitch-tracking analyses were performed using routines coded in Matlab 7.4.1 (2005; The MathWorks, Natick, MA). For the stimulus-to-response correlation measure, analysis was conducted separately

for the two segments (0–139 ms, 139–278 ms) as per our division of the stimulus into steady-state and dynamic portions (Fig. 1).

**Auditory (sound to meaning) training.** Overall, the participant's task was to learn 24 words (6 pseudowords  $\times$  4 lexical tones) produced by the 4 different talkers over 9 training sessions. Procedures were identical to those from a previous study (Chandrasekaran et al. 2010). In each training session, the participant learned to pair objects with sounds (e.g., /peshT2/ = pencil; /peshT3/ = clock). Each training session had the following sequence of events: participants listened to each sound paired with the visual presentation of the object. After the auditory-visual pairing, a short test was conducted to test the participant's sound-to-object association skills. Feedback was provided to facilitate learning correct sound-object pairings. At the end of the training session, a final test was conducted during which no feedback was provided. In this test participants heard a word (e.g., /peshT2/) and had to pick one of 24 objects that referred to that word. In total, participants heard 96 sounds ( $24 \times 4$  talkers). A final score (proportion word identification) was calculated from each of the nine sessions. The word identification score from the ninth session (S9) was used as an index of learning success.

**Behavioral tests.** Participants were also evaluated with a nonverbal IQ test (Test of Nonverbal Intelligence; Brown et al. 1997) as well as measures of phonological awareness and auditory working memory using subtests of the Woodcock-Johnson Test of Cognitive Abilities (Woodcock et al. 2001).

## RESULTS

**Pitch pattern representation in IC RS and IC RE groups.** Participants were divided into two groups (IC RS and IC RE, see Fig. 3) based on IC fMRI signal, extracted from the IC ROI (Fig. 2) in the tone change > repeat contrast. Considerable variability was seen in the direction of repetition effects (Fig. 3). A number of participants ( $n = 10$ , IC RS) showed positive activity in this contrast (tone change > repeat), whereas others ( $n = 8$ , IC RE) showed greater activity in the repeat condition relative to the tone-change condition. Crucially, the two groups did not differ on the basis of cognitive scores and were comparable in terms of nonverbal IQ scores, phonological processing (performance on sound-blending task), and auditory working memory (Table 1). As a control, participants were also divided into two groups on the basis of activity in a second contrast (speaker change > repeat). The speaker-change condition involved changing the talker (change in average pitch but not in pitch direction) but not the tone per se. In this control grouping, some participants ( $n = 11$ ) showed response suppression (IC speaker-RS) whereas others ( $n = 7$ ) showed response enhancement (IC speaker-RE). Group differences in pretraining IC pitch pattern representation, as evidenced by FFRs, were statistically evaluated. Two FFR indexes were examined: FFR amplitude, which reflects the number of frequency-following neural ensembles activated by the stimulus, and stimulus-to-response correlation, which reflects similarity

(Pearson's moment-to-moment correlation) between pitch extracted (using autocorrelation methods) from the stimulus to that of the response. For the stimulus-to-response correlation index, we divided the stimulus into two equal portions that reflect the complexity of pitch movement within the syllable (steady state, dynamic).

**FFR amplitude measure.** First, we examined differences in pretraining FFR amplitude between the IC RS and IC RE groups. We then conducted a second ANOVA on the FFR prestimulus amplitude to examine whether group differences could be explained by general differences in background EEG noise not related to the stimulus. Finally, as a control, we examined differences in FFR amplitude between groups divided on the basis of speaker-change condition (IC speaker-RS, IC speaker-RE). ANOVA conducted on FFR amplitude revealed a significant main effect of group [IC RS vs. IC RE:  $F(1,16) = 8.67$ ,  $P = 0.01$ ]. Means and standard deviations revealed significantly larger FFR amplitude in the RE group relative to the RS group (RE > RS, Fig. 4), indicative of more efficient neural processing (i.e., fewer neural ensembles encoding the stimulus cycles) in the RS group. No group differences were found in prestimulus amplitude [ $F(1,16) = 1.89$ ,  $P = 0.19$ ], demonstrating that group FFR amplitude differences are stimulus related. ANOVA conducted to examine FFR amplitude differences on classification based on the control condition (speaker change > repeat) did not reveal a main effect of group [ $F(1, 16) = 1.23$ ,  $P = 0.27$ ].

**FFR stimulus-to-response correlation.** The stimulus-to-response correlation measure, reflective of the fidelity of the FFR, was entered into a 2 (steady-state vs. dynamic portion of T2)  $\times$  2 (RE vs. RS) rmANOVA, revealing a significant segment  $\times$  group interaction effect [ $F(1,16) = 5.06$ ,  $P = 0.039$ ], a main effect of segment [ $F(1,16) = 37.84$ ,  $P < 0.001$ ], and no main effect of group [ $F(1,16) = 1.79$ ,  $P = 0.2$ ]. One-way ANOVA revealed that individuals in the IC RS group, relative to the IC RE group, showed more faithful representation of the stimulus in the dynamic portion of the response [ $F(1,16) = 8.06$ ,  $P = 0.012$ ] but not in the steady-state portion [ $F(1,16) = 0.33$ ,  $P = 0.58$ ; Fig. 4]. Means indicate that individuals who showed IC RS had a positive stimulus-to-response correlation and those with IC RE showed negative stimulus-to-response correlation during the dynamic portion of the response. Negative correlation is indicative of inability to track the rapidly rising portion of the stimulus (pitch errors in the opposite direction of pitch change, Fig. 4). Consistent with our predictions, the group division based on the control condition (speaker change > repeat) did not reveal any significant main effects of group [ $F(1,16) = 0.12$ ,  $P = 0.73$ ] or interaction effects [ $F(1,16) = 0.02$ ,  $P = 0.88$ ].

In conclusion, we examined group differences in stimulus transcription, as revealed by the FFR. We found that the IC RS group, relative to the IC RE group, showed smaller FFR amplitude and better representation of the stimulus pitch pattern in the dynamic portion of the stimulus. Together, these are indicative of greater neural efficiency as well as enhanced IC stimulus representation in the RS group relative to the RE group.

**Pretraining fMRI IC repetition effects relate to auditory learning success.** We examined the extent to which individual differences in IC pitch encoding related to auditory learning success. After the pretraining fMRI and EEG recordings,

Table 1. Cognitive scores for IC RS and IC RE groups

	IC RS	IC RE	P Value (2-tailed)
Nonverbal IQ	116 (2.9)	117 (3.5)	0.86
Phonological awareness	82 (4.6)	78 (7.8)	0.67
Auditory working memory	0.88 (2.1)	0.81 (4.1)	0.17

Values are means and SE (in parentheses) for participants who showed repetition suppression (RS) or repetition enhancement (RE) indexed by the inferior colliculus (IC) functional MRI signal.

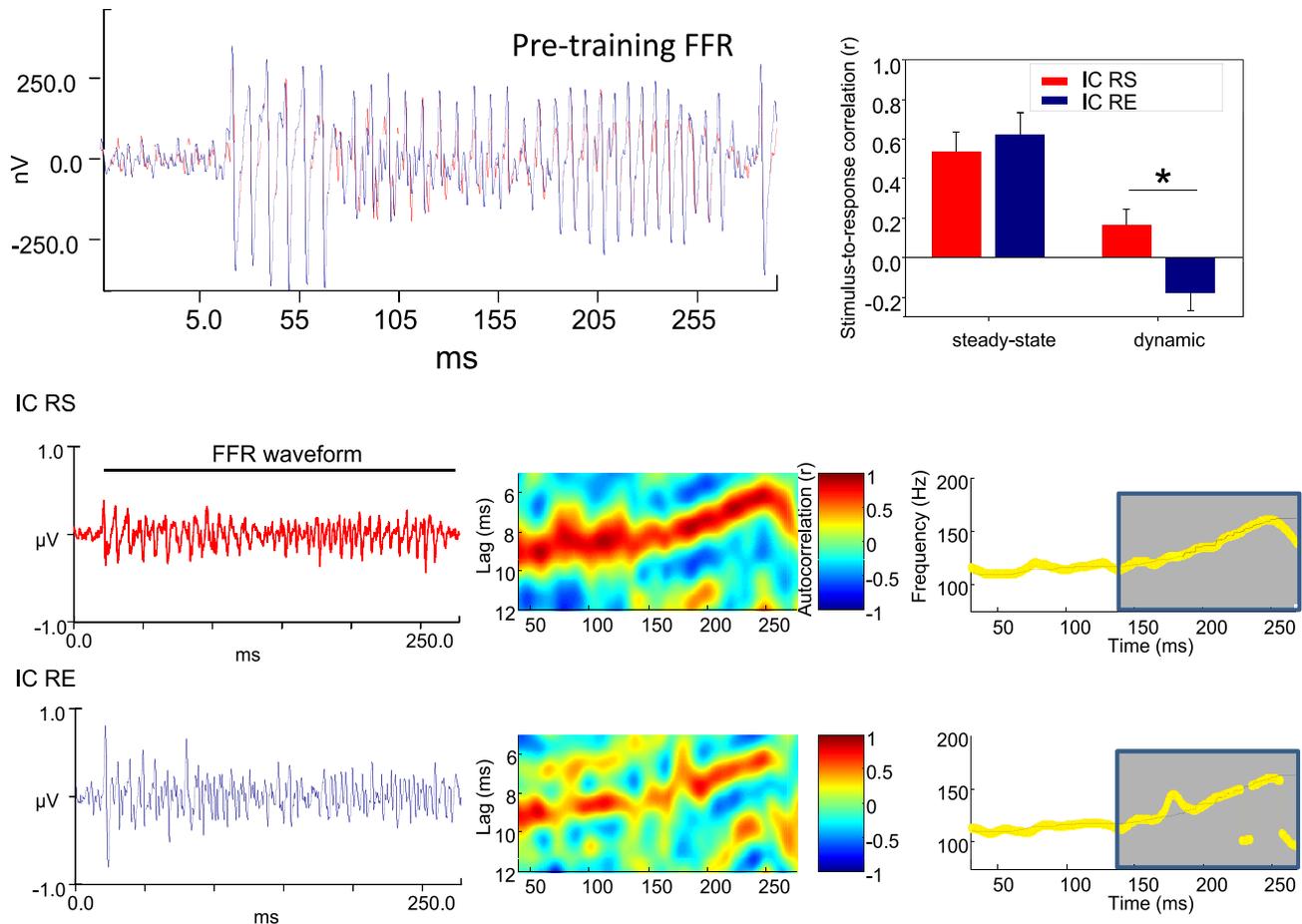


Fig. 4. *Top left*: group differences in IC pitch representation as reflected by the FFR. *Top right*: red and blue bars show mean stimulus-to-response correlation from individuals in the IC RS and IC RE groups (see Fig. 3), respectively (error bars indicate SE). Both groups show poorer response fidelity in the dynamic portion of the response to the stimulus, relative to the steady-state portion. However, group differences are only seen in the dynamic portion (Fig. 1) of the FFR ( $P = 0.012$  based on independent samples  $t$ -tests). *Bottom left*: FFRs of a representative participant in the IC RS group (red) relative to a participant in the IC RE group (blue). *Bottom right*: FFR analysis using autocorrelation methods; the extracted pitch contour is compared (correlated) with the stimulus pitch contour. The participant in the IC RS group shows reduced FFR amplitude but greater fidelity (higher stimulus-to-response correlation) relative to the individual in the IC RE group. These data show that individual differences in IC fMRI signal relate to differences in encoding the complexity of the stimulus.

participants underwent a 9-day sound-to-meaning training program during which they learned to use pitch patterns to discern words (Fig. 5). RmANOVA conducted on word identification scores collected during each session revealed a significant session  $\times$  group interaction effect [ $F(8,9) = 6.48, P = 0.006$ ]. One-way ANOVA revealed that individuals in the IC RS group had significantly better word learning scores relative to the IC RE group, from the third session (S3) onwards. Group differences persisted through the last session (Fig. 5). S1 and S2 did not differentiate the two groups [S1:  $F(8,9) = 4.073, P = 0.06$ ; S2:  $F(8,9) = 1.248, P = 0.28$ ]. From S6–S9, the two groups showed significantly different word identification scores [S6:  $F(8,9) = 12.1, P = 0.003$ ; S7:  $F(8,9) = 11.2, P = 0.004$ ; S8:  $F(8,9) = 8.82, P = 0.009$ ; S9:  $F(8,9) = 10.07, P = 0.006$ ]. These data strongly demonstrate learning differences between the two groups, suggesting that pretraining IC representation of pitch patterns is predictive of auditory learning success.

Group division based on the control condition (IC speaker-RS vs. IC speaker-RE) did not reveal any significant main effects of group [ $F(1,16) = 2, P = 0.18$ ] or interaction effects [ $F(8,9) = 0.779, P = 0.63$ ].

*Training modulates FFR representation of stimulus pitch pattern.* We compared ABRs collected pretraining and posttraining to the same pitch pattern (T2) to examine training-related changes in stimulus representation. RmANOVA conducted on FFR amplitude revealed a significant training (pretraining, posttraining)  $\times$  group (IC RS, IC RE) interaction effect [ $F(1,16) = 13.22, P = 0.002$ ] and marginally significant main effects of training [ $F(1,16) = 4.44, P = 0.051$ ] and group [ $F(1,16) = 3.34, P = 0.09$ ]. Post hoc  $t$ -tests revealed a significant effect of training (posttraining < pretraining) for the RE group ( $P = 0.03$ ) but not for the RS group ( $P = 0.66$ ). Posttraining, FFR amplitude reduced for the RE group, reflecting more efficient neural encoding after nine sessions of sound-to-meaning training (Fig. 6). RmANOVA were conducted on the stimulus-to-response correlation measure separately for the steady-state and dynamic portions of the stimulus. RmANOVA conducted on the dynamic portion revealed a significant main effect of training [ $F(1,16) = 6.87, P = 0.019$ ] and group [ $F(1,16) = 5.74, P = 0.03$ ] but no significant interaction effect. In contrast, no main effect of training [ $F(1,16) = 0.001, P = 0.97$ ] or group [ $F(1,16) = 0.06, P = 0.81$ ] or interaction effect [ $F(1,16) = 0.51, P = 0.48$ ] were

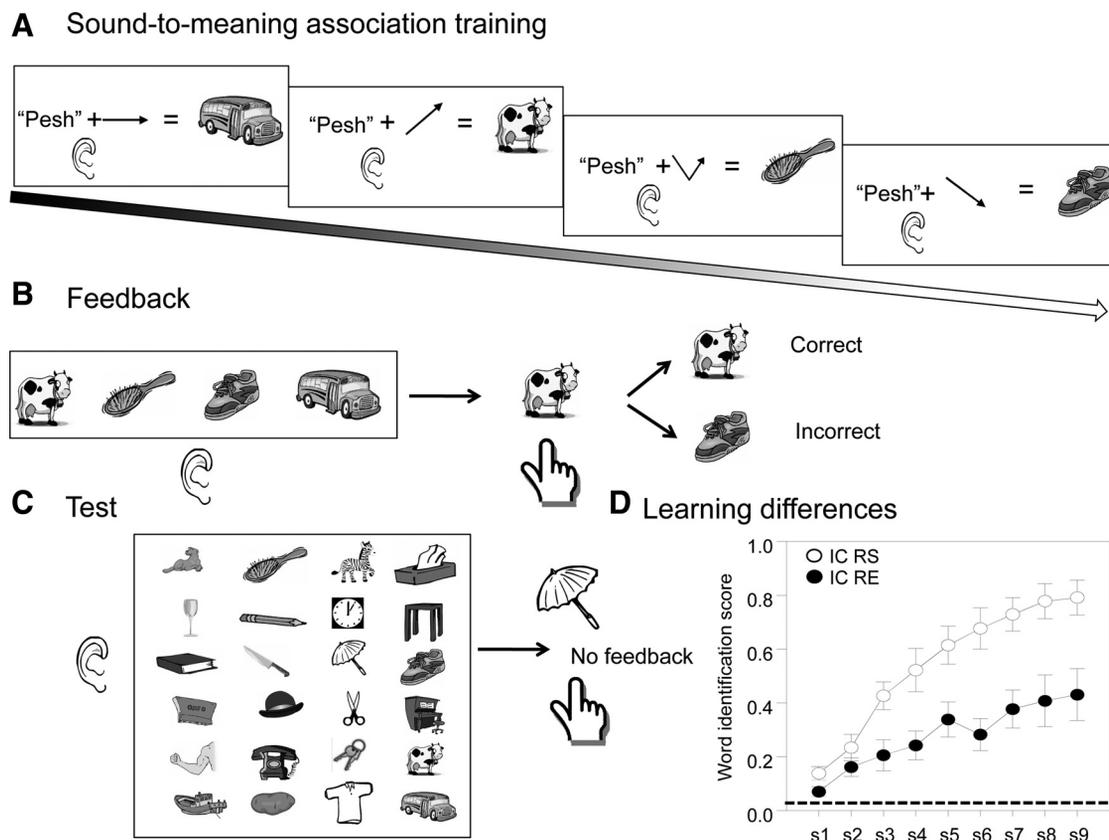


Fig. 5. Participants underwent a 9-session sound-to-meaning association training paradigm during which they learned to associate pseudowords (e.g., “Pesh”) with superimposed pitch patterns (level, rising, dipping, or falling pitch patterns, indicated by arrows) to differentiate objects. Each training session had the following sequence of events. **A:** sounds ( $n = 24$ , 4 tones  $\times$  6 pseudowords) were paired with visual presentation of a unique object (e.g., “Pesh” spoken with a rising pitch pattern is associated with “cow,” and with a falling pitch pattern is associated with “shoe”). **B:** after the auditory-visual pairing, a feedback test was conducted to investigate the participant’s sound-to-object association. During this test, participants heard a sound and had to choose 1 of 4 objects on the screen. Feedback [correct/incorrect (with the correct object)] was provided to facilitate learning correct sound-object pairings. **C:** each session ended with a final test during which no feedback was provided. In this test, participants heard a sound (e.g., /peshT4/) and had to pick 1 of 24 objects. A final score (proportion word identification) was calculated from each of the 9 sessions. **D:** proportion of words correctly identified for the 2 groups (IC RS, IC RE) across 9 sessions of training showing clear separation after the second session (S2). The IC RS group, relative to the IC RE group, was more successful in the sound-to-meaning training program, suggesting that IC encoding of nonnative time-varying pitch patterns is related to successful linguistic pitch pattern learning. Word identification scores were significantly above chance ( $\sim 0.04$ , as indicated by dotted line) for both groups.

found in the steady-state portion, suggesting that training did not change the neural representation of steady-state pitch. Means and standard deviations indicate that training improved stimulus-to-response correlations for both groups (posttraining  $>$  pretraining, Fig. 6). Irrespective of training, the IC RS group showed superior encoding of dynamic pitch relative to IC RE (Fig. 6). Together, these data show that training modulates the representation of dynamic pitch patterns. With training, the IC RE group showed a reduction in FFR amplitude, consistent with the proposal that a sparser representation is associated with more enhanced stimulus representation. Irrespective of group, pitch representation in the steady-state segment was robust and did not change with training. In contrast, pitch representation in the dynamic portion of the segment was poor before training and improved after training (increase in stimulus-to-response correlation). Both groups showed improved stimulus-to-response correlations posttraining, but only in the dynamic portion of the stimulus. Overall, the IC RS group showed superior dynamic pitch representation relative to the IC RE group (pretraining as well as posttraining). Together, these data reveal that training modulates early sensory representation of dynamic pitch patterns.

The group division based on the control condition (IC speaker-RS, IC speaker-RE) did not reveal any significant main effects of training [ $F(1,16) = 0.26$ ,  $P = 0.61$ ] or group [ $F(1,16) = 0.01$ ,  $P = 0.92$ ] or any interaction effects [ $F(1,16) = 3.05$ ,  $P = 0.1$ ] for the FFR amplitude measure or stimulus-to-response measure [dynamic: main effect of group (IC speaker-RS vs. IC speaker-RE):  $F(1,16) = 0.17$ ,  $P = 0.67$ ; group  $\times$  training interaction effect:  $F(1,16) = 0.002$ ,  $P = 0.965$ ].

## DISCUSSION

We examined the extent to which the IC fMRI signal, as indexed by fMRI-adaptation, relates to auditory learning success. The use of multimodal imaging affords the means to evaluate the relationship between the direction of repetition effects (RS vs. RE) and the quality of stimulus representation in the IC. Based on IC neural repetition effects, participants were divided into two groups, those who showed RS and those who showed RE (Fig. 3). We examined differences in the neurophysiological representation of pitch between groups, using FFRs as an index. FFRs reflect IC transcription of the stimulus with millisecond precision (Fig. 1). Before training,

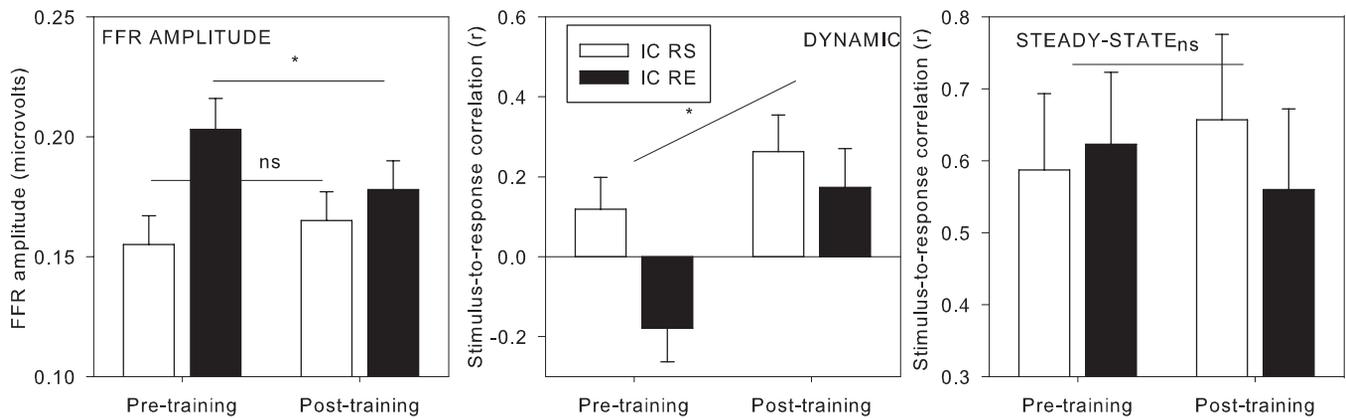


Fig. 6. Effects of sound-to-meaning training on FFR indexes. *Left*: training improved the neural efficiency of brain stem pitch pattern encoding in the IC RE group. Before training, the IC RE group showed larger FFR amplitude to T2 relative to the IC RS group. Training reduced FFR amplitude for the IC RE group; no differences were seen for the IC RS group. *Middle* and *right*: training selectively improved the IC representation of dynamic pitch. Stimulus-to-response correlations were calculated for the dynamic (*middle*) and the steady-state (*right*) portions of the stimulus pitch pattern. FFRs recorded posttraining showed greater fidelity to the dynamic pitch contour relative to the pretraining session for both groups. In contrast, no training-related changes were found in the response to the steady-state portion of the stimulus for either group. Values are means  $\pm$  SE. \* $P < 0.05$ . ns, Not significant at  $P < 0.05$ .

we found that individuals who showed RS had smaller FFR amplitudes to the rising (T2) pitch pattern relative to those who showed RE, suggesting greater neural efficiency in stimulus representation (Fig. 4). Importantly, the RS group also showed enhanced representation of this stimulus dimension relative to the RE group. As indexed by stimulus-to-response correlations, individuals in the RS group showed more faithful stimulus pitch tracking to the rising portion of the stimulus before training, relative to the RE group (Fig. 4). The two groups showed equivalent representation of steady-state pitch patterns, suggesting that group differences were specific to neural encoding of the portion of the stimulus that involved dynamic pitch change. Together, these results show that individuals who showed RS to pitch patterns, relative to those who exhibited RE, showed more efficient (sparser) and enhanced neural representation of specific features (dynamic pitch) within the stimulus.

Next, we examined the extent to which pretraining IC repetition effects to pitch patterns predicted differences in learning to use these pitch patterns to contrast words in the sound-to-meaning training paradigm. Overall, participants who showed IC RS were more successful in word learning across the nine sessions of training (Fig. 5). We also divided participants into two groups (IC speaker-RS, IC speaker-RE) based on the control contrast (speaker change > repeat). In the speaker-change condition, the pitch pattern was the same across all trials, whereas the speaker changed. Although this contrast (speaker change > repeat) also involved pitch differentiation, the difference is based on pitch height (i.e., average pitch) rather than pitch direction. We found no representational differences (as indexed by FFR) between the two groups (IC speaker-suppression), no learning differences, and no training-related neuroplasticity. We take this as evidence that the fMRI adaptation method is highly sensitive in examining neurophysiological signal encoding differences to specific dimensions (pitch direction vs. pitch height) within the signal. Our results demonstrate that pretraining midbrain sensory encoding of pitch patterns is associated with the ability to use nonnative pitch patterns in lexical contexts.

Importantly, neurophysiological encoding differences are not static in adults. Our results demonstrate that training

improves the representation of dynamic pitch change, demonstrating plasticity even in the mature auditory brain stem. These results are consistent with a previous study that showed training-related effects on the FFRs (Song et al. 2008). Song et al. (2008) used a similar sound-to-meaning paradigm and demonstrated pitch pattern encoding differences posttraining relative to pretraining. Although Song et al. (2008) did not find training-related effects on the encoding of the rising pitch contour (T2), this could be attributed to several reasons; individual differences were not evaluated in this study, and interestingly, FFR indexes to T2 showed the most intersubject variability, suggesting significant individual differences in encoding accuracy. In the current study, participants were divided into two different groups to account for individual differences in IC encoding. Furthermore, we separately evaluated encoding for the steady-state portion (that showed minimal individual differences and high fidelity even before training) and the dynamic portion (that showed significant individual differences). Posttraining, both groups showed improved representation of the dynamic portion of the pitch pattern, as reflected by greater correlation between the stimulus pitch pattern and the response pitch pattern (Fig. 6). Consistent with our predictions, no training effects were seen in the response to the steady-state portion of the stimulus. Since we used a multital-ker (high variability) sound-to-meaning training paradigm, participants had to focus on pitch direction as a cue (because speaker-related cues are irrelevant for discerning words). Consequently, posttraining, we saw highly specific training effects on neurophysiological encoding of dynamic pitch change. Our results are consistent with previous studies that have shown highly specific training-related effects on the FFRs (Carcagno and Plack 2011). Interestingly, posttraining, the IC RE group showed a significant reduction in FFR amplitude, suggesting improved neural efficiency (sparser encoding). Training did not modulate FFR amplitude in the IC RS group, suggesting that these individuals were already representing the signal in an efficient manner even before training.

In adults, perceiving nonnative phonemes is one of the biggest challenges to learning a foreign language (Iverson et al. 2003). A major challenge in learning tone languages is difficulty in using pitch to distinguish words (Wang et al. 1999).

Our results reveal a neurobiological basis for individual variability in word-learning ability. Since pitch is a multidimensional attribute, pitch patterns can be distinguished using multiple cues. Previous cross-language studies on tone perception have shown that native speakers of a tone language tend to emphasize pitch direction as cue more than nonnative speakers of a tone language (Chandrasekaran et al. 2007; Gandour and Harshman 1978). Even in nonnative speakers, individuals who tend to attend more to pitch direction as a cue are more likely to be successful in learning to associate pitch patterns with words (Chandrasekaran et al. 2010). In animal models, the IC nuclei have been shown to contain neurons specialized in processing the direction of FM, a cue that is used in animal vocalizations (Andoni et al. 2007; Brimijoin and O'Neill 2005). Since individuals who show IC RS show better encoding of the dynamic portion of the stimulus (during which pitch is changing rapidly), these participants may be better able to distinguish pitch patterns and consequently learn to associate the complex pitch patterns with words. Importantly, the two groups did not differ on the basis of cognitive or phonological ability, ruling out nonsensory differences between groups. Neuroimaging studies examining individual differences in human auditory learning have thus far focused on the cortex (Golestani et al. 2007; Norton et al. 2005; Warrier et al. 2009; Wong et al. 2007a). Our results show that individual differences in auditory learning success are also related to midbrain sensory representation, demonstrating that the operational specifics of human auditory learning cannot be completely understood by exclusively focusing on cortical structures (Parvizi 2009).

Several studies have reported substantial variability in second language phonetic learning (Chandrasekaran et al. 2010; Iverson et al. 2003; Wong et al. 2011). The exact reasons for such individual differences are unknown, although several reasons have been put forth. A previous study showed that part of the individual differences in phonetic categorization arise from acoustic sensitivities to irrelevant acoustic cues (Iverson et al. 2003). Specifically, in this study, Japanese listeners who showed poor phonetic categorization differences between English /r/ and /l/ also exhibited greater acoustic sensitivity to second formant cues (that does not differentiate /r/ and /l/) and less reliance on the third formant (a cue that differentiates /r/ and /l/). In the current study, individual differences in phonetic (tone) learning may arise from differential sensitivity to pitch height relative to pitch direction (a cue that differentiates Mandarin pitch contours). The ultimate reasons for these midbrain differences may arise from natural variation resulting from experiential or genetic factors. Previous studies have shown that musical training may be an important experiential factor that results in superior brain stem encoding of speech sounds (Kraus and Chandrasekaran 2010; Musacchia et al. 2007; Wong et al. 2007). In the current study, we specifically controlled for this factor, yet we found significant individual differences in learning success. Genetic factors underlying individual differences in sensory encoding have not been systematically evaluated. A recent study showed that the allele frequency of two genes, *ASPM* and *MCPHI*, are associated with the presence of linguistic tone (Dediu and Ladd 2007). The authors hypothesized that these genes may play a causal role in determining interindividual differences in lexical tone learning. With respect to the results from the current study,

further work is needed to determine the relative contribution of experiential vs. genetic factors in determining individual differences in midbrain processing of important linguistic cues.

Our findings also provide new insights into understanding subcortical neural repetition effects. Previous studies have shown that neuronal responses in the mammalian auditory cortex are highly sensitive to stimulus repetition (Bartlett and Wang 2005; Deouell et al. 2007; Ulanovsky et al. 2003, 2004). With repeated presentations of a stimulus, cortical neural activity is reduced in some regions (Desimone 1996; Grill-Spector et al. 2006; Henson and Rugg 2003; Krekelberg et al. 2006). Since this reduction (RS) is often associated with behavioral improvements (repetition priming), the reduction in cortical neural activity is argued to reflect increased neural efficiency. As per the fatigue model, stimulus repetition results in a reduction in firing rates in neuronal pools sensitive to the stimulus features that are repeated. In contrast, in the sharpening model, repetition results in pruning of neurons that are not relevant in encoding the stimulus; with repetition, neurons become more finely tuned to stimulus properties (Desimone 1996; Gagnepain et al. 2008; Grill-Spector et al. 2006). Finally, facilitation models have argued that the BOLD reduction with repetition reflects faster neural processing, associated with top-down-driven fulfillment of perceptual expectations. None of these models specifically explain individual differences in the direction of neural repetition effects. Repetition enhancement, as per fatigue models, results from an increase in neural firing rate; this cannot explain why individuals who show RE had poorer signal representation. Furthermore, since the perceptual expectations were the same for all participants, the facilitation model, which views RS effects as resulting from validated top-down expectation (Summerfield et al. 2008), cannot explain individual differences in neural repetition effects. Of the various models, the sharpening model can satisfactorily explain our results. As per the sharpening model, neurons functionally adapt to repetitive stimuli by fine-tuning to relevant features; neurons that code features irrelevant for signal representation are pruned out with repetition, resulting in activation of fewer neurons. Although this results in neural response suppression, the representation of key features is enhanced (or fine-tuned) with repetition (Grill-Spector et al. 2006). In the current study, individuals who showed IC RS showed reduced FFR amplitude, suggesting that fewer neural phase-locking ensembles are activated by the stimulus in these individuals. These participants also showed superior encoding of crucial signal features (dynamic pitch). In contrast, individuals who showed IC RE showed increased FFR amplitude, suggesting more neuronal ensembles activated by the stimulus. Since these participants showed poorer encoding of crucial signal features (e.g., pitch direction), we hypothesize that repetition enhancement likely occurs due to the recruitment of additional neuronal pools that are needed to encode the complexity of the dynamic changes in pitch pattern. RE could also occur due to an increased metabolic cost in encoding complexity. These results show that individual differences in the direction of repetition effects relate to neurophysiological differences in stimulus representation, which in turn may lead to behavioral differences in the ability to use pitch direction cues meaningfully.

Previous studies have used a multimodal approach [EEG or magnetoencephalography (MEG) in combination with fMRI]

to examine the physiological basis of repetition-induced response reduction at the level of the cortex (Gonsalves et al. 2005; McDonald et al. 2010; Vidyasagar et al. 2010). Because fMRI has relatively poor temporal resolution and EEG/MEG has relatively poor spatial resolution, examining both forms of information from the same participant affords a better insight into neurophysiology of repetition effects. However, relative to examining cortical evoked potentials, brain stem electrophysiology offers unique advantages to understanding the physiological basis of the brain's response to repetition. The FFR, unlike cortical responses, provide more direct information about the degree to which neural activity reflects the acoustic properties of the stimulus (Chandrasekaran and Kraus 2010; Kraus and Chandrasekaran 2010; Skoe and Kraus 2010). Not only can we discern reduction in FFR amplitude, we can directly compare the response to the stimulus (stimulus-to-response correlation), providing a clearer index of stimulus-related activity. Furthermore, the correspondence between IC fMRI signal and FFRs in our data speaks to the extent to which the fMRI signal reflects synaptic activity vs. spiking activity (Goense and Logothetis 2008; Logothetis 2007; Viswanathan and Freeman 2007). The cycle-to-cycle phase-locking by the FFR is argued to reflect synaptic (excitatory postsynaptic potential) activity, rather than action potentials (Britt 1976). The strong correspondence between the IC BOLD response and FFR provide further evidence that the fMRI signal reflects synaptic activity.

In conclusion, our results highlight the importance of the IC in pitch pattern encoding as well as in auditory learning in humans. We anticipate our findings using a combined fMRI-A/brain stem electrophysiology approach to be a starting point for a more sophisticated understanding of auditory midbrain processing to speech patterns and the dynamics of human auditory learning.

## GRANTS

This work was supported by National Science Foundation Grant BCS-0719666 and National Institute on Deafness and Other Communication Disorders Grant R01 DC008333.

## DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

## AUTHOR CONTRIBUTIONS

Author contributions: B.C., N.K., and P.W. conception and design of research; B.C. performed experiments; B.C. analyzed data; B.C., N.K., and P.W. interpreted results of experiments; B.C. prepared figures; B.C. drafted manuscript; B.C., N.K., and P.W. edited and revised manuscript; B.C., N.K., and P.W. approved final version of manuscript.

## REFERENCES

- Andoni S, Li N, Pollak GD.** Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. *J Neurosci* 27: 4882–4893, 2007.
- Bajo VM, Nodal FR, Moore DR, King AJ.** The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci* 13: 253–260, 2010.
- Bartlett EL, Wang X.** Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol* 94: 83–104, 2005.
- Baumann S, Griffiths TD, Sun L, Petkov CI, Thiele A, Rees A.** Orthogonal representation of sound dimensions in the primate midbrain. *Nat Neurosci* 14: 423–425, 2011.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B.** Voice-selective areas in human auditory cortex. *Nature* 403: 309–312, 2000.
- Brimijoin WO, O'Neill WE.** On the prediction of sweep rate and directional selectivity for FM sounds from two-tone interactions in the inferior colliculus. *Hear Res* 210: 63–79, 2005.
- Britt RH.** Intracellular study of synaptic events related to phase-locking responses of cat cochlear nucleus cells to low frequency tones. *Brain Res* 112: 313–327, 1976.
- Brown L, Sherbenou RJ, Johnsen SK.** *TON13: Test of Nonverbal Intelligence* (3rd ed.). Austin, TX: Pro-Ed, 1997.
- Carcagno S, Plack CJ.** Subcortical plasticity following perceptual learning in a pitch discrimination task. *J Assoc Res Otolaryngol* 12: 89–100, 2011.
- Chandrasekaran B, Chan AH, Wong PC.** Neural processing of what and who information in speech. *J Cogn Neurosci* 23: 2690–2700, 2011.
- Chandrasekaran B, Gandour JT, Krishnan A.** Neuroplasticity in the processing of pitch dimensions: a multidimensional scaling analysis of the mismatch negativity. *Restor Neurol Neurosci* 25: 195–210, 2007.
- Chandrasekaran B, Kraus N.** The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47: 236–246, 2010.
- Chandrasekaran B, Sampath PD, Wong PC.** Individual variability in cue-weighting and lexical tone learning. *J Acoust Soc Am* 128: 456–465, 2010.
- Church BA, Schacter DL.** Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *J Exp Psychol Learn Mem Cogn* 20: 521–533, 1994.
- Cox RW.** AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29: 162–173, 1996.
- Deidu D, Ladd DR.** Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proc Natl Acad Sci USA* 104: 10944–10949, 2007.
- Deouell LY, Heller AS, Malach R, D'Esposito M, Knight RT.** Cerebral Responses to Change in Spatial Location of Unattended Sounds. *Neuron* 55: 985–996, 2007.
- Desimone R.** Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci USA* 93: 13494–13499, 1996.
- Gagnepain P, Chetelat G, Landeau B, Dayan J, Eustache F, Lebreton K.** Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. *J Neurosci* 28: 5281–5289, 2008.
- Galbraith GC, Bagasan B, Sulahian J.** Brainstem frequency-following response recorded from one vertical and three horizontal electrode derivations. *Percept Mot Skills* 92: 99–106, 2001.
- Gandour JT, Harshman RA.** Crosslanguage differences in tone perception: a multidimensional scaling investigation. *Lang Speech* 21: 1–33, 1978.
- Goense JB, Logothetis NK.** Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr Biol* 18: 631–640, 2008.
- Golestani N, Molko N, Dehaene S, LeBihan D, Pallier C.** Brain structure predicts the learning of foreign speech sounds. *Cereb Cortex* 17: 575–582, 2007.
- Gonsalves BD, Kahn I, Curran T, Norman KA, Wagner AD.** Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. *Neuron* 47: 751–761, 2005.
- Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD.** Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci* 4: 633–637, 2001.
- Grill-Spector K.** Selectivity of adaptation in single units: implications for fMRI experiments. *Neuron* 49: 170–171, 2006.
- Grill-Spector K, Henson R, Martin A.** Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10: 14–23, 2006.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW.** “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7: 213–223, 1999.
- Henson R, Shallice T, Dolan R.** Neuroimaging evidence for dissociable forms of repetition priming. *Science* 287: 1269–1272, 2000.
- Henson RN.** Neuroimaging studies of priming. *Prog Neurobiol* 70: 53–81, 2003.
- Henson RN, Rugg MD.** Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia* 41: 263–270, 2003.
- Horner AJ, Henson RN.** Priming, response learning and repetition suppression. *Neuropsychologia* 46: 1979–1991, 2008.
- Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, Siebert C.** A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57, 2003.

- Jäncke L, Gaab N, Wustenberg T, Scheich H, Heinze HJ.** Short-term functional plasticity in the human auditory cortex: an fMRI study. *Brain Res Cogn Brain Res* 12: 479–485, 2001.
- Kraus N, Chandrasekaran B.** Music training for the development of auditory skills. *Nat Rev Neurosci* 11: 599–605, 2010.
- Krekelberg B, Boynton GM, van Wezel RJ.** Adaptation: from single cells to BOLD signals. *Trends Neurosci* 29: 250–256, 2006.
- Krishnan A, Swaminathan J, Gandour JT.** Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci* 21: 1092–1105, 2009.
- Logothetis NK.** The ins and outs of fMRI signals. *Nat Neurosci* 10: 1230–1232, 2007.
- McDonald CR, Thesen T, Carlson C, Blumberg M, Girard HM, Trong-trung A, Sherfey JS, Devinsky O, Kuzniecky R, Dolye WK, Cash SS, Leonard MK, Hagler DJ Jr, Dale AM, Halgren E.** Multimodal imaging of repetition priming: using fMRI, MEG, and intracranial EEG to reveal spatiotemporal profiles of word processing. *Neuroimage* 53: 707–717, 2010.
- Musacchia G, Sams M, Skoe E, Kraus N.** Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 104: 15894–15898, 2007.
- Norton A, Winner E, Cronin K, Overy K, Lee DJ, Schlaug G.** Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain Cogn* 59: 124–134, 2005.
- Parvizi J.** Corticocentric myopia: old bias in new cognitive sciences. *Trends Cogn Sci* 13: 354–359, 2009.
- Peach RK, Wong PC.** Integrating the message level into treatment for agrammatism using story retelling. *Aphasiology* 18: 429–441, 2004.
- Perrachione TK, Wong PC.** Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. *Neuropsychologia* 45: 1899–1910, 2007.
- Schreiner CE, Langner G.** Laminar fine structure of frequency organization in auditory midbrain. *Nature* 388: 383–386, 1997.
- Schreiner CE, Langner G.** Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60: 1823–1840, 1988.
- Skoe E, Kraus N.** Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31: 302–324, 2010.
- Smith JC, Marsh JT, Brown WS.** Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 39: 465–472, 1975.
- Smith JC, Marsh JT, Greenberg S, Brown WS.** Human auditory frequency-following responses to a missing fundamental. *Science* 201: 639–641, 1978.
- Song JH, Skoe E, Banai K, Kraus N.** Training to improve hearing speech in noise: biological mechanisms. *Cereb Cortex*. First published July 28, 2011. doi:10.1093/cercor/bhr196.
- Song JH, Skoe E, Wong PC, Kraus N.** Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 20: 1892–1902, 2008.
- Strait DL, Kraus N, Skoe E, Ashley R.** Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci* 29: 661–668, 2009.
- Suga N.** Role of corticofugal feedback in hearing. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 194: 169–183, 2008.
- Suga N, Ma X.** Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat Rev Neurosci* 4: 783–794, 2003.
- Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T.** Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci* 11: 1004–1006, 2008.
- Sun X, Xia Q, Lai CH, Shum DK, Chan YS, He J.** Corticofugal modulation of acoustically induced Fos expression in the rat auditory pathway. *J Comp Neurol* 501: 509–525, 2007.
- Turk-Browne NB, Yi DJ, Leber AB, Chun MM.** Visual quality determines the direction of neural repetition effects. *Cereb Cortex* 17: 425–433, 2007.
- Ulanovsky N, Las L, Farkas D, Nelken I.** Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci* 24: 10440–10453, 2004.
- Ulanovsky N, Las L, Nelken I.** Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6: 391–398, 2003.
- Vidyasagar R, Stancak A, Parkes LM.** A multimodal brain imaging study of repetition suppression in the human visual cortex. *Neuroimage* 49: 1612–1621, 2010.
- Viswanathan A, Freeman RD.** Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nat Neurosci* 10: 1308–1312, 2007.
- von Kriegstein K, Patterson RD, Griffiths TD.** Task-dependent modulation of medial geniculate body is behaviorally relevant for speech recognition. *Curr Biol* 18: 1855–1859, 2008.
- Wang Y, Sereno JA, Jongman A, Hirsch J.** fMRI evidence for cortical modification during learning of Mandarin lexical tone. *J Cogn Neurosci* 15: 1019–1027, 2003b.
- Wang Y, Spence MM, Jongman A, Sereno JA.** Training American listeners to perceive Mandarin tones. *J Acoust Soc Am* 106: 3649–3658, 1999.
- Warrier CM, Wong PC, Penhune V, Zatorre R, Parrish T, Abrams D, Kraus N.** Relating structure to function: Heschl's gyrus and acoustic processing. *J Neurosci* 29: 61–69, 2009.
- Williams AJ, Fuzessery ZM.** Facilitatory mechanisms shape selectivity for the rate and direction of FM sweeps in the inferior colliculus of the pallid bat. *J Neurophysiol* 104: 1456–1471, 2010.
- Wong FCK, Chandrasekaran B, Garibaldi K, Wong PCM.** White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *J Neurosci* 31: 8780–8785, 2011.
- Wong PC, Perrachione TK, Parrish TB.** Neural characteristics of successful and less successful speech and word learning in adults. *Hum Brain Mapp* 28: 995–1006, 2007a.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N.** Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10: 420–422, 2007b.
- Woodcock RW, McGrew KS, Mather N.** Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside, 2001.
- Zhang LI, Tan AYY, Schreiner CE, Merzenich MM.** Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* 424: 201–205, 2003.