Published in final edited form as:

J Cogn Neurosci. 2008 October ; 20(10): 1892–1902. doi:10.1162/jocn.2008.20131.

Plasticity in the adult human auditory brainstem following short-term linguistic training

Judy H. Song[1,3], **Erika Skoe**[1,3], **Patrick C. M. Wong**[2,3,4,6], and **Nina Kraus**[1,3,4,5,6]

[1]Auditory Neuroscience Laboratory, Northwestern University

[2]Communication Neural Systems Research Group, Northwestern University

[3]The Roxelyn and Richard Pepper Department of Communication Sciences & Disorders, Northwestern University

[4]Northwestern University Institute for Neuroscience, Northwestern University

[5]Department of Neurobiology and Physiology, Northwestern University

[6]Department of Otolaryngology, Northwestern University

Abstract

Peripheral and central structures along the auditory pathway contribute to speech processing and learning. However, because speech requires the use of functionally and acoustically complex sounds which necessitates high sensory and cognitive demands, long-term exposure and experience using these sounds is often attributed to the neocortex with little emphasis placed on subcortical structures. The present study examines changes in the auditory brainstem, specifically the frequency following response (FFR), as native English-speaking adults learn to incorporate foreign speech sounds (lexical pitch patterns) in word identification. The FFR presumably originates from the auditory midbrain, and can be elicited pre-attentively. We measured FFRs to the trained pitch patterns before and after training. Measures of pitch-tracking were then derived from the FFR signals. We found increased accuracy in pitch-tracking after training, including a decrease in the number of pitch-tracking errors and a refinement in the energy devoted to encoding pitch. Most interestingly, this change in pitch-tracking accuracy only occurred in the most acoustically complex pitch contour (dipping contour), which is also the least familiar to our English-speaking subjects. These results not only demonstrate the contribution of the brainstem in language learning and its plasticity in adulthood, but they also demonstrate the specificity of this contribution (i.e., changes in encoding only occurs in specific, least familiar stimuli, not all stimuli). Our findings complement existing data showing cortical changes after second language learning, and are consistent with models suggesting that brainstem changes resulting from perceptual learning are most apparent when acuity in encoding is most needed.

Keywords

auditory learning; brainstem encoding; speech perception; tone language

Introduction

Before entering the neocortex, successive neural impulses initiated by sounds entering the cochlea reach subcortical structures, including the brainstem. As acoustic and/or functional

Corresponding Author: Patrick Wong, Ph.D., Communication Sciences & Disorders, 2240 Campus Drive, Evanston, IL, (847) 491-2416 (phone), (847) 491-2429 (fax), pwong@northwestern.edu.

complexity of sound increases, the more likely higher-level structures will be involved in processing (Gordon and O'Neill 2000; Suga et al. 2000). Specifically, the processing and plasticity of the encoding of speech sounds is generally attributed to the neocortex (Liebenthal et al. 2005; Zatorre et al. 1992), although brainstem and thalamic structures contribute to such processing to a certain extent. Evidence of such high-level (including cognitive) involvement and learning-associated plasticity comes from studies of long-term and short-term perceptual learning of speech (Kraus et al. 1995; Naatanen et al. 1997; Tervaniemi et al. 2006; Tremblay et al. 1997).

There is a growing body of evidence to suggest that subcortical structures contribute actively to auditory processing, and are not simply passive relay stations transmitting information from the peripheral sensory organs to the cortex. Recent studies have shown that the auditory brainstem can be modified by short-term and long-term auditory experiences initiated in childhood. For example, Russo et al. (2005) found improved auditory brainstem timing to speech stimuli in background noise in children with language-based learning problems following an eight-week commercially available auditory speech training program. Krishnan et al. (2005) measured the impact of long-term language experience on the frequency following response (FFR). They found that native Mandarin-speaking subjects with at least twenty years of Mandarin language exposure (beginning in childhood) showed more precise linguistic pitch pattern encoding relative to native English-speaking subjects. Mandarin Chinese, a tone language, uses pitch, the psychological correlate of F0, to signal word meaning at the syllable level (e.g., /ma/ spoken with high-level and rising pitch patterns mean ‘mother’ and ‘numb’, respectively)[1]. This increased neural precision reflects Mandarin-speakers' long-term learning of Mandarin tones (lexically meaningful pitch patterns), and how this experience has changed the response properties of subcortical neurons for enhanced processing of linguistic pitch patterns. Similarly, long-term musical training initiated in childhood has been shown to enhance frequency encoding in the brainstem such that English-speaking musicians who did not speak Mandarin showed more robust and faithful encoding of Mandarin tones, especially the dipping (i.e., falling then rising) contour (Wong et al. 2007b). Enhancements due to life-long musical training have also been found in the brainstem encoding of acoustically transient events and harmonic content for both speech and music (Musacchia et al. 2007) It is important to note that all of these studies involve auditory experiences initiated in childhood. It is yet to be demonstrated whether short-term experiences occurring in adulthood can have any measurable impact on the brainstem.

The present study examines changes in the auditory brainstem, specifically the FFR, as native English speaking adults learn to incorporate foreign speech sounds (lexical pitch patterns) in word identification. The FFR, a far field potential recorded from surface electrodes, reflects the synchronized activities of axonal and dendritic potentials generated by populations of neurons in the lateral lemniscus and/or inferior colliculus of the brainstem (Hoormann et al. 1992; Smith et al. 1975) and is well suited for examining how speech-specific pitch contours are encoded subcortically. There is a vast literature demonstrating the existence of a temporal code of pitch encoding at the level of auditory nerve and brainstem (Langner 1997; Moller 1999). This temporal code is observed in discharge patterns of single neurons, and in synchronous population-wide neuronal activity. The pattern of neural discharge (phase-locking) is modulated by the temporal structure of the eliciting sound. In the temporal structure of speech sounds, periodic amplitude modulations reflect the rate of the F0, and elicit the perception of pitch. This periodicity is maintained in the neuronal representation, with interspike intervals entraining to the period of the F0 and its harmonics. The acoustic features of the evoking stimulus, both spectral and temporal, are represented with high fidelity in the

[1] The use of pitch to distinguish word meaning in a tone language is similar to the use of consonants to contrast word meaning in English (e.g., ‘pet’ and ‘bet’ only differs in the initial consonant and resulted in two different words in English).

FFR, making it possible to compare the response frequency composition and timing to the corresponding features of the stimulus (Galbraith et al. 2000; Hall 1979; Johnson et al. 2005; Kraus and Nicol 2005).

We trained native-English speaking adults to use three different pitch patterns (or tones): high-level, rising, and dipping pitch patterns in word identification. For example, subjects learned that the pseudoword “pesh” spoken with a high-level, rising, and dipping tone meant ‘glass’, ‘pencil’, and ‘table’, respectively. These tonal patterns resemble those used lexically in Mandarin Chinese. High-level and rising tones are acoustically simpler than dipping tones (Fig. 1) and are used frequently in English as intonational (non-lexical) markers at the syllable level. The dipping tone, which contains a large downward then upward excursion, can only occur at the phrase level in English (Pierrehumbert 1979) and is the most difficult tone for second language learners to master (Gottfried and Suiter 1997;Kiriloff 1969).

The FFR was elicited pre-attentively to the three training tones superimposed onto an untrained syllable (/mi/). Tone 3, which was the least familiar to the subjects, served as the experimental stimulus, while Tone 1 and Tone 2, which occur in English, served as within-subject control stimuli. Subjects' neural pitch-tracking accuracy to the stimulus F0 was assessed before and after they were trained on our lexical pitch task. If short-term linguistic learning can result in brainstem plasticity even in adulthood, we would expect pitch-tracking accuracy to improve. This improvement would be most evident in the dipping tone which is the most acoustically complex (as its shape involves a falling-rising pattern) and is the least familiar to the subjects.

Materials and Methods

Subjects

Twenty-three subjects (14 females) native English-speaking adults (age: 19-40 years; mean age: 26.3 ± 5 years) participated in this study. All subjects reported no audiologic or neurologic deficits and had normal click-evoked auditory brainstem response latencies with right ear stimuli presentation at 80 dB SPL (Hood 1998). Hearing thresholds were screened at 20 dB HL for octaves from 500 to 4000 Hz for both ears. All subjects had normal IQ (mean IQ: 119 ± 13.6), as measured by Wechsler's Abbreviated Scale of Intelligence (WASI) (Wechsler 1999). Informed consent was obtained from all subjects. The Institutional Review Board of Northwestern University approved this research study.

Training

Stimuli—Mandarin is a tonal language that utilizes changes in pitch to indicate word meaning. In this study, a male native speaker of Mandarin Chinese produced the syllable /mi/ with three Mandarin tones: high-level (Tone 1), rising (Tone 2), and dipping (Tone 3) (Fig. 1). The F0 contours of the training and electrophysiological stimuli were both created based on these three speech tokens. For the training study, the F0 contours from the 3 syllables were extracted (Fig. 1) and superimposed onto 6 English pseudowords, to form 6 minimal triads using the Pitch-Synchronous Overlap and Add (PSOLA) resynthesis method implemented and documented in Praat (Boersman and Weeknick 2005). The pseudowords follow English phonotactic rules words but are not part of the English lexicon. The 6 monosyllabic pseudowords (i.e., “dree,” “fute,” “ner,” “nuck,” “pesh,” and “vece”; formal phonetic transcriptions are provided in Table I) were produced by a male native speaker of American English in a sound attenuated chamber via a SHURE SM58 microphone onto a Pentium IV PC sampled at 44.1 kHz. English pseudowords were used because unknown words containing native phonological patterns are easier to learn than those with non-native phonological patterns (Feldman and Healy 1998). The F0 contour was duration normalized to fit the length of each originally produced pseudoword. In other words, although the pitch trajectory was identical in the spectral domain for all pseudowords using the same tone (i.e. “pesh1”, “dree1”, “ner1”, etc.), the rate of frequency modulation (how frequency change over time) was different depending on the duration of the originally produced pseudoword. All stimuli were amplitude normalized such that after resynthesis, each originally produced pseudoword had three variants that differed only in F0 with the duration, syllable onset, rhyme and coda being identical. These 18 (6 × 3) resynthesized stimuli, with pitch contours from a male Mandarin speaker and syllables from an American English speaker, were used in the training program. Six native Mandarin-speaking adults judged the training stimuli to be perceptually natural; each identified the pitch patterns of the training stimuli with at least 95% accuracy.

Procedures—The training program lasted eight sessions. Each session, including the training blocks, practice quizzes, and test, lasted approximately 30 minutes. All subjects completed training in 14 consecutive days, with no more than two days between sessions. These training stimuli and procedures were adapted from our previous study (Wong and Perrachione 2007; Wong et al. 2007a). Subjects were trained to associate pseudowords with drawings that represented high frequency English nouns. In order to facilitate learning, the 18 pseudowords were split into minimal contrast triads (six groups of three stimuli). The training session was divided into six blocks, and the subject was trained on one triad per block, similar to our previous study. The order of the blocks was randomized across training sessions. Training involved the simultaneous presentation of the sound of the pseudoword via headphones and corresponding picture. Within each block, subjects were presented each sound-picture pair four times resulting in a total of 72 (6 blocks × 4 times × 3 pitch contours) pseudoword-picture presentations during each daily training session. At the end of each block, subjects were given a practice quiz which required them to match the pseudoword with one of three drawings. Subjects received feedback on their performance -- if the correct picture was identified, “Correct” was displayed on the computer screen, and “Incorrect. The correct answer is … <correct picture shown>” was displayed if the wrong picture was chosen. After one block was completed, the next block began immediately and this was repeated until all six blocks were completed.

After completing the six blocks, subjects were tested on the entire set of 18 pseudowords without feedback. This test presented each pseudoword one at a time, randomized and repeated three times (total of 54 trials). Subjects were instructed to identify each word by selecting the corresponding drawing out of 18 possible choices. Subjects were allowed to take as much time as needed to associate word and picture. The word identification score was obtained at the end of each session in order to monitor the subjects' progress.

Physiologic Responses to Pitch Patterns (Tones)

Physiologic Stimuli—The same three tokens of the syllable /mi/ containing the three Mandarin tones [high-level (Tone 1), rising (Tone 2) and dipping (Tone 3)] were also used to generate stimuli for the FFR. These utterances were duration-normalized to 278.5 ms, resulting in stimuli which contained the same pitch trajectories in the spectral domain as the training stimuli but differed in the rate of frequency modulation. F0 contours from each syllable were then extracted and superimposed onto the original /mi1/ syllable, following resynthesis procedures described above. This resulted in three stimuli which, other than differing in F0, were acoustically identical and were judged to be perceptually natural by four native speakers of Mandarin. In Mandarin, the syllable /mi/ spoken with these tones translates ‘to squint’, ‘to bewilder’, ‘rice’, respectively. The minimum and maximum frequencies of F0 contours of the three stimuli were 140-172 Hz, 110-163 Hz, and 89-110 Hz, respectively (Fig. 1). These stimuli were RMS amplitude normalized using the software Level 16 (Trice and Carrell 1998). To accommodate the capabilities of our stimulus presentation software, the stimuli were resampled to 22.05 kHz. These stimuli are identical to those in our previous study (Wong et al. 2007b) and were not used in training.

Physiologic Recording Procedures—During the pre- and post-training sessions, subjects watched a videotape with the sound level set at less than 40 dB SPL to facilitate a quiet yet wakeful state. Subjects' left ears were unoccluded to allow for the delivery of the video soundtrack, while the stimuli were presented to the right ear at ~70 dB SPL (Neuroscan Stim; Compumedics, El Paso, TX) through insert ear phones (ER-3; Etymotic Research, Elk Grove Village, Ill). The order of the three stimuli was randomized across subjects with a variable inter-stimulus interval between 71.50 and 104.84 ms. Responses were collected using Scan 4.3 (Neuroscan; Compumedics, El Paso, TX) with three Ag–AgCl scalp electrodes, differentially recorded from Cz (active) to ipsilateral earlobe (reference), with the forehead as ground. Contact impedance was less than 5 kΩ for all electrodes. Two blocks of 1200 sweeps per block were collected at each polarity with a sampling rate of 20 kHz. Filtering, artifact rejection and averaging were performed offline using Scan 4.3. Responses were bandpass filtered from 80 to1000 Hz, 12 dB/octave, and trials with activity greater than ±35 μV were considered artifacts and rejected. Waveforms were averaged with a time window spanning 45 ms prior to the onset and 16.5 ms after the offset of the stimulus. Responses of alternating polarity were then added together to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Gorga et al. 1985). For the purpose of calculating signal-to-noise ratios, a single waveform representing non-stimulus-evoked neural activity was created by averaging the neural activity 40 ms prior to stimulus onset.

Analysis Procedures—In order to assess training-induced physiologic changes, we measured subjects' FFRs elicited by the three trained pitch patterns (tones) embedded in the untrained syllable /mi/ (Fig. 1 shows the F0 contours of these stimuli). Physiologic data were collected immediately before the first session of training and immediately after the last session of training. For each subject, three measures of FFR pitch tracking were calculated: *Pitch Tracking Error*, *Spectral Dominance of F0*, and *Pitch Noise Ratio*, which were used to assess the subjects' pitch tracking to the stimulus F0 contours of the three stimuli. These measures were derived using a sliding window analysis procedure, in which 40-ms bins of the FFR were analyzed in the frequency domain. The FFR was assumed to encompass the entire response beginning at time 1.1 ms, the transmission delay between the ER-3 transducer and ear insert. The 40-ms sliding window was shifted in 1-ms steps to produce a total of 238 overlapping FFR bins. A narrow-band spectrogram was calculated for each Hanning-windowed FFR bin by applying the Fast Fourier Transform (FFT). To increase spectral resolution, each time bin was zero-padded to 1 second before performing the FFT. The spectrogram gave an estimate of spectral energy over time and the F0 (pitch) contour was extracted from the spectrogram by finding the spectral peak closest to the expected (stimulus) frequency. F0 frequency and amplitude were recorded for each time bin. The same short-term spectral analysis procedure was applied to the stimulus waveforms. In Fig. 4 and 5, the time indicated on the *x*-axis refers to the midpoint of each 40-ms time bin analyzed. *Pitch Tracking Error* was calculated using both linear and logarithmic scales. *Linear Pitch Tracking Error*, a measure of pitch encoding accuracy over the duration of the stimulus, was calculated by finding the absolute Euclidian distance between the stimulus F0 and response F0 at each time bin and averaging the error across all 238 bins. In computing *Logarithmic Pitch Tracking Error*, the stimulus F0 and the response F0 were transformed to log units before finding the average absolute difference between the stimulus and the response.

Spectral Dominance of F0 and *Pitch Noise Ratio* are two measures of spectral amplitude, which consider the extent to which the extracted F0s meet or surpass a specific threshold across the entire stimulus. Specifically, *Spectral Dominance of F0* describes whether the extracted F0s were at the spectral maximum and reflects the strength of F0 encoding for the duration of the stimulus in the brainstem. This measure of spectral amplitude was calculated by finding the number of time bins in which the extracted F0 fell at the spectral maximum (largest peak in spectrum). For each time bin, a signal-to-noise ratio (SNR) was also calculated. This was done by applying a FFT to the 40-ms waveform representing the non-stimulus-evoked activity (*noise*) and then finding the spectral amplitude corresponding to F0 that was extracted from the respective FFR bin. SNRs were calculated as $F0Amplitude_{FFR\ BINx} / F0Amplitude_{NOISE}$, where x is a number from 1 to 238, and F0 is the frequency extracted from bin x. *Pitch Noise Ratio* describes whether the extracted F0s were above the noise floor and thus, represents the number of bins for which the SNR was greater than one. It reflects the magnitude of the response in encoding stimulus F0 relative to the ongoing neural response. All pitch-tracking analyses were performed using routines coded in Matlab 7.0.4 (The Mathworks, Natick, MA).

Results

Behavioral Measures (Tone Training Program)

Subjects' word identification performance was evaluated at the end of each training session. We found that immediately after the first training session, subjects' mean word identification was 21.56% (range 4.17 to 58.33%). At the end of the last training session (henceforth “attainment”), subjects' mean performance was 89.49% (range 11.11 to 100%), an improvement of 67.93%, which is statistically significant as revealed by a paired t-test [t = -17.06, $p < 0.0001$]. Fig. 2 shows subjects' mean learning trajectory.

Physiologic Responses (Pitch Pattern/Tone Stimuli)

We hypothesized that learning-induced brainstem modifications would only occur for Tone 3 (dipping tone), thus we first report results concerning Tone 3. For Tone 3 *Linear Pitch Tracking Error*, a one-way repeated measures ANOVA revealed a main effect of training [F (1, 22) = 14.343, p = 0.001] (Fig. 3A). Thus, subjects' brainstem response exhibited more faithful representation of the dipping stimulus F0 contour after being trained to use this tone in a lexical context. Fig. 4 shows examples of pitch tracking of the brainstem response to Tone 3 before and after training from three representative subjects (panel A) as well as word identification scores from their first and last training sessions (panel B). The same pattern of results was observed for *Logarithmic Pitch Tracking Error* [F (1, 22) = 12.897, p = 0.002]. Moreover, not only was the F0 of Tone 3 encoded more precisely after training, the manner in which it was encoded also changed. After training, F0 was more likely to be encoded by the largest spectral peak as indicated by an increase in *Spectral Dominance* in the post-training data. A one-way repeated measures ANOVA on Tone 3 S*pectral Dominance* showed a main effect of training [F (1, 22) = 4.878, *p* = 0.038]. Likewise, *Pitch Noise Ratio* also increased with training which resulted in fewer points below the noise floor. A one-way repeated measures ANOVA on Tone 3 *Pitch Noise Ratio* also showed a main effect of training [F (1, 22) = 4.454, *p* = 0.046]. Fig. 5 shows a representative subject's FFR waveforms, FFR spectra and pitch tracking for the three tones after training, along with the corresponding values for *Linear Pitch Tracking Error, Spectral Dominance,* and *Pitch Noise Ratio*[2].

As predicted, FFR responses to Tones 1 and 2, the control pitch stimuli did not show measurable changes after training (Fig. 3A). One-way repeated measures ANOVAs on Tones 1 and 2 showed no significant changes after training for *Linear Pitch Tracking Error*, [F (1, 22) = 0.014, *p* = 0.907, F (1, 22) = 0.477, *p* = 0.497, respectively], *Spectral Dominance*, [F (1, 22)

[2] We also performed correlational analyses between behavioral performance (accuracy in word identification post-training) and the various physiologic measures. When all subjects were combined, no significant correlations were found. However, seven subjects performed at ceiling behaviorally (100% accuracy in word identification). When those seven subjects were excluded, post-training behavioral performance was significantly correlated with post-training Tone 3 *Linear Pitch Tracking Error* [Spearman's rho = .55 (*p* = .014)] and *Log Pitch Tracking Error* [Spearman's rho = .61 (*p* = .006)].

= 0.034, *p* = 0.855, F (1, 22) = 0.126, *p* = 0.726, respectively], and *Pitch Noise Ratio* [F (1, 22) = 0.001, *p* = 0.980, F (1, 22) = 0.165, *p* = 0.688, respectively]. In addition to the absence of mean differences, changes in brainstem responses for Tones 1 and 2 were also more variable compared to Tone 3 (Fig. 3B-C). For most of the subjects, the percentage of pitch-tracking errors decreased after training for Tone 3, whereas for Tones 1 and 2, no consistent pattern of change was observed. The fact that brainstem improvement of pitch tracking occurred only with Tone 3 suggests the impact of short-term linguistic training on subcortical circuitry is highly specific and most evident in the aspect of speech (in this case, tonal pattern) that is least familiar to the learners.

Discussion

Our results demonstrate plasticity in the adult human auditory brainstem following short-term linguistic training. We measured changes in auditory brainstem encoding of variable F0 (pitch) patterns by examining the FFR, a subcortical response presumably originating from the rostral brainstem that encodes the F0 (a physiological correlate of perceived pitch) of the stimulus with high fidelity (Marsh and Worden 1968). We found that after learning to use three pitch patterns for word identification at the syllable level, native-English speaking adults showed increased accuracy in pitch tracking. It is important to point out that these native-English speaking adults had no experience using pitch lexically prior to our training program. This increase was revealed by a decrease in the number of pitch-tracking errors and a refinement in the energy devoted to encoding pitch, including increased signal-to-noise ratios and increased *Spectral Dominance* of the stimulus pitch. Although our subjects had experience using high-level and rising pitch contours at the syllable level for signaling intonation in English, they had less experience with the dipping tone. We found that pitch-tracking improvements occurred only in this least familiar pitch pattern.

While Krishnan et al. found that native Mandarin speakers had increased accuracy in pitch tracking compared to native English speaking adults (Krishnan et al. 2005), and Wong et al. (2007b) and Musacchia et al. (2007) found enhanced brainstem encoding of the F0 in musicians, these studies can only speak to the effect of long-term auditory experiences initiated in childhood. Moreover, while short-term training has been shown to improve brainstem timing in children with learning problems (Russo et al. 2005), these findings had not yet been extended to adults. To the best of our knowledge, we are the first to show that experiences acquired in adulthood can modify brainstem auditory responses in a context specific way.

Our study adds to a growing body of research focusing on the neural encoding of linguistic pitch contours. Taken together, findings from these studies are consistent with a fundamental operating principle of brainstem function, specifically that pitch is represented in a temporal code (phase-locking). This is underscored by the fact that our subjects showed combined improvements in the behavioral and neurophysiological representation of pitch likely reflecting *enhanced* synchronization of neuronal firing to the stimulus F0. This enhancement may be the result of additional neurons firing at the rate of the stimulus F0, the same population of neurons firing more synchronously, or a combination of the two. Moreover, learning may also engender the synchronization or enhancement of populations of neurons which synchronize to the F0 as evidenced by EEG (Bao et al. 2004; Musacchia et al. 2007; Shahin et al. 2003; Tremblay et al. 2001) and MEG studies (Fujioka et al. 2006; Shahin et al. 2007) both measures reflect the summation of electrical activity generated by synchronous firing of neurons. In addition to the strengthening of the temporal code, our results may reflect an increase in accuracy of representation via the place code, as the frequencies of our stimuli were low enough in frequency to be associated with both temporal and place code.

Our findings can be interpreted within the framework of corticofugal tuning. While we know that top-down control from the auditory cortex to peripheral auditory receptors can occur in human adults via electrical stimulation of the cortex and through selective attention on perceptual tasks, active training-induced corticofugal modulation has not been shown. In the present study, our subjects received spoken language training through the mapping of new sound structures and lexical-semantic concepts. Their substantial behavioral improvements suggest that they were actively engaged in linguistic learning. The training program involved both high cognitive demands (as it is a case of spoken language learning) and auditory acuity. Thus, it not only engaged the neocortex, but also subcortical structures, respectively. It is conceivable that feedback from higher-level cortex is initiated so that precise pitch information can be relayed to the neocortex to successfully perform the cognitively demanding task. This comports with models of perceptual learning involving changes in the weighting of perceptual dimensions as a result of feedback (Nosofsky 1986). Applied to our current study, these models would suggest that the attentional weighting of pitch-relevant dimensions increased as the result of a training task requiring attention to pitch. Perceptual weighting adjustments were also observed in a recent animal study which showed that visual learning in the presence of irrelevant auditory signals dampened sensitivity to the auditory signals (Delano et al. 2007). This line of reasoning is also consistent with the Reverse Hierarchy Theory (RHT) of visual learning suggesting that learning consists of an attention-driven, task-dependent “backward” search for increased signal-to-noise ratio, especially in highly-skilled performers and perceptual experts (Ahissar and Hochstein 2004).

Relevant to RHT we found brainstem plasticity to be associated with the most acoustically complex tone, which happened to also be the tone that was least familiar to the subjects. Although our subjects did not have experience using any pitch patterns lexically, as native English speakers, they did have experience using high-level and rising tones at the syllable level for contrasting intonational meaning. As such, they have experience encoding these two pitch patterns. However, the dipping tone is only used at the phrase level in English (e.g., “You should go home now, MY DEAR FRIEND [the last three words are said in a dipping contour]”). Therefore, our learners had little experience encoding this tone as rapidly as required at the syllable level. Per RHT, lower-level neural involvements are especially pronounced in more experienced learners and in conditions in which encoding acuity is most needed. Our learners as a group reached above 89% accuracy (most above 90%) and can be argued to be quite experienced on this task. Furthermore, the dipping tone, being most acoustically complex with its falling-rising pattern, arguably most requires the acoustic acuity. Thus, it is not surprising that improvements in brainstem encoding occurred only with this tone. The finding that training effects were only seen for this most complex and previously less familiar tone also suggest a complexity threshold that the cortex needs in order to drive and reinforce subcortical tuning. This complexity threshold relates to both linguistic (Tone 3 being the least familiar tone linguistically) and acoustic (Tone 3 having the most complex pitch contour) complexity. Additional experiments are needed to determine whether both linguistic and acoustic complexities are needed to produce the largest tuning effect in the brainstem.

Interestingly, in our recent study using the same stimulus set, we examined the pitch-tracking accuracy of native English-speaking adult musicians and nonmusicians who had no previous exposure to a tone language. We found that the two groups were best differentiated by their tracking of this dipping tone. More specifically, musicians exhibited better tracking to Tone 3 (but not Tone 1) than nonmusicians (Wong et al. 2007b) suggesting that long-term musical experience may give particular advantage to brainstem encoding of complex linguistic pitch patterns. This advantage occurred despite that fact that these particular pitch patterns were not linked specifically to musical training. Note that this seemingly context-general tuning effect (music affecting speech) could be attributed to auditory experience initiated in childhood given that our musician subjects began their musical training early in life. In contrast, it is possible that for experiences initiated in adulthood, such as those in the present study, that brainstem tuning is entirely context-specific and therefore restricted to the (novel) training stimuli. Furthermore, the length of learning can influence the degree of retention. Retention may be impeded following short-term training in adults due to the strong activation of established neural circuitry hampering the acquisition of alternative patterns of circuit connectivity (Knudsen 2004). Thus in order to compete with entrenched circuitry and facilitate plasticity, effective attentional shifts and repeated presentations of the content are required (Bahrick and Hall 2004; Lively et al. 1993).

It is worth noting that we are not ruling out possible passive, bottom-up learning to account for the results. Passive exposure, simple perceptual learning, and sensitivity to stimulus statistical distributions have been found to be associated with behavioral improvements (Maye et al. 2002; Seitz and Watanabe 2003; Watanabe et al. 2001) and brainstem encoding (Dean et al. 2005; Escabi et al. 2003). Furthermore, neurons in the rostral brainstem have been found to be sensitive to pitch trajectories independent of functional contexts (Gordon and O'Neill 2000). However, similar to Ahissar and Hochstein, we believe top-down influence to dominate the lower-level (in our case, brainstem) responses given evidence suggesting that practice and active performance, rather than passive learning or generalization, tend to dominate neural changes (Recanzone et al. 1992). All subjects in the current study were required to actively perform the task and all showed substantial behavioral improvements. It is unlikely that passive exposure to the sounds would result in the same effect we observed (Schoups et al. 2001), especially given that only the least familiar pitch contour showed pitch-tracking improvements. It is worth pointing out that Tone 2 (rising tone) was the tone that showed the most pitch tracking errors before training. The lack of improvement in this tone, therefore, cannot be attributed to a pre-training ceiling effect. This finding suggests that perceptual learning is an unlikely candidate for explaining the patterns of results we observed.

An important aspect of our results is that although improvement in encoding only occurred with one tone (Tone 3), behavioral improvements demonstrated by word identification were substantial. Our learning paradigm focuses on the matching of eighteen auditory stimuli to eighteen pictures. The correct use of segmental (consonants and vowels) information alone will likely account for some of the improvement. Furthermore, strengthening the sensitivity of one category in a three-category perceptual space can result in all three categories being more distinguishable as demonstrated by models of perceptual learning [see (Fahle and Poggio 2002) for a review]. We believe that in fact, this is what we observed in the current study. Furthermore, as discussed, the fact that only Tone 3, the tone that is least familiar to the subjects in its linguistic (high-level) usage, resulted in brainstem encoding changes, provides strong evidence for a top-down explanation, although a simple perceptual learning mechanism cannot be ruled out without further experimentation.

We have established that brainstem modifications can occur after short-term training even when it is initiated in adulthood. It is worth noting that although the speech training literature at large has shown that the most efficacious training paradigm (measured by subjects' ability to generalize to new talkers) is one that uses variable stimuli (Clopper and Pisoni 2004; Lively et al. 1993), a low-variability program (with training stimuli spoken by one talker) was used. As the aim of the current study was to establish, for the first time, a training-induced brainstem effect in adults, we have chosen to adopt a simpler training paradigm. Future work could include high-variability paradigms and examine their effect on brainstem plasticity. Future work could also consider brainstem plasticity resulting from different types of complex auditory learning (e.g., music) initiated both in adulthood and childhood to examine any possible quantitative and qualitative differences. This work would inevitably have an impact on education and clinical (re)habilitation (e.g. to inform strategies for improving auditory processing in the rapidly growing population of hearing impaired adults) and would assist in informing social and educational policies.

Acknowledgments

The authors wish to thank Ann Bradlow, Trent Nicol, and Tyler Perrachione for their assistance in this research. This work is supported by Northwestern University and the National Institutes of Health grants R03HD051827 and R21DC007468 to P.W., and R01DC001510 to N.K., as well as National Science Foundation grant BCS-0544846 to N.K.

References

Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 2004;8:457–464. [PubMed: 15450510]

Bahrick H, Hall L. The important of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. J Mem Lang 2004;52:566–577.

Bao S, Chang E, Woods J, Merzenich M. Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. Nat Neurosci 2004;7:974–881. [PubMed: 15286790]

Boersman, P.; Weeknick, D. PRAAT: Doing phonetics by computers. 2005. http://www.fon.hum.uva.nl/praat/, v. 4.3.04

Clopper CG, Pisoni DB. Effects of talker variability on perceptual learning dialects. Lang Speech 2004;47:207–239. [PubMed: 15697151]

Dean I, Harper NS, McAlpine D. Neural population coding of sound level adapts to stimulus statistics. Nat Neurosci 2005;8:1684–1689. [PubMed: 16286934]

Delano PH, Elgueda D, Hamame CM, Robles L. Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. J Neurosci 2007;27:4146–4153. [PubMed: 17428992]

Escabi MA, Miller LM, Read HL, Schreiner CE. Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. J Neurosci 2003;23:11489–11504. [PubMed: 14684853]

Fahle, M.; Poggio, T. Perceptual Learning. The MIT Press; 2002.

Feldman, A.; Healy, AF. Foreign Language Learning: Psychological Studies on Training & Retention. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.; 1998.

Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ. One year of musical training affects development of auditory cortical-evoked fields in young children. Brain 2006;129:2593–2608. [PubMed: 16959812]

Galbraith GC, Threadgill MR, Hemsley J, Salour K, Songdej N, Ton J, Cheung L. Putative measure of peripheral and brainstem frequency-following in humans. Neurosci Lett 2000;292:123–127. [PubMed: 10998564]

Gordon M, O'Neill WE. An extralemniscal component of the mustached bat inferior colliculus selective for direction and rate of linear frequency modulations. J Comp Neurol 2000;426:165–181. [PubMed: 10982461]

Gorga, M.; Abbas, P.; Worthington, D. Stimulus calibration in ABR measurements. In: Jacobson, J., editor. The Auditory Brainstem Response. San Diego: College-Hill; 1985.

Gottfried TL, Suiter TL. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. J Phonetics 1997;25:207–231.

Hall JW 3rd. Auditory brainstem frequency following responses to waveform envelope periodicity. Science 1979;205:1297–1299. [PubMed: 472748]

Hood, L. Clinical Applications of the Auditory Brainstem REsponse. San Diego: Singular Publishing Group, Inc.; 1998.

Hoormann J, Falkenstein M, Hohnsbein J, Blanke L. The human frequency-following response (FFR): normal variability and relation to the click-evoked brainstem response. Hear Res 1992;59:179–188. [PubMed: 1618709]

Johnson KL, Nicol T, Kraus N. The brainstem response to speech: a biological marker of auditory processing. Ear Hear 2005;26:424–434. [PubMed: 16230893]

Kiriloff C. On the auditory perception of tones in Mandarin. Phonetica 1969;20:63–67.

Knudsen E. Sensitive periods in the development of the brain and behavior. J Cogn Neurosci 2004;16:1412–1425. [PubMed: 15509387]

Kraus N, McGee T, Carrell TD, King C, Tremblay K, Nicol T. Central auditory system plasticity associated with speech discrimination training. J Cogn Neurosci 1995;7:25–32.

Kraus N, Nicol T. Brainstem origins for cortical ‘what’ and ‘where’ pathways in the auditory system. Trends Neurosci 2005;28:176–181. [PubMed: 15808351]

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

Langner G. Neural processing and representation of periodicity pitch. Acta Otolaryngol Suppl 1997;532:68–76. [PubMed: 9442847]

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cereb Cortex 2005;15:1621–1631. [PubMed: 15703256]

Lively SE, Logan JS, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. J Acoust Soc Am 1993;93:1242–1255. [PubMed: 8408964]

Marsh JT, Worden FG. Sound evoked frequency-following responses in the central auditory pathway. Laryngoscope 1968;78:1149–1163. [PubMed: 5659588]

Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 2002;82:B101–111. [PubMed: 11747867]

Moller AR. Review of the roles of temporal and place coding of frequency in speech discrimination. Acta Otolaryngol 1999;119:424–430. [PubMed: 10445056]

Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci U S A 2007;104:15894–15898. [PubMed: 17898180]

Naatanen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 1997;385:432–434. [PubMed: 9009189]

Nosofsky RM. Attention, similarity, and the identification-categorization relationship. J Exp Psychol Gen 1986;115:39–61. [PubMed: 2937873]

Pierrehumbert J. The perception of fundamental frequency declination. J Acoust Soc Am 1979;66:363–369. [PubMed: 512199]

Recanzone GH, Jenkins WM, Hradek GT, Merzenich MM. Progressive improvement in discriminative abilities in adult owl monkeys performing a tactile frequency discrimination task. J Neurophysiol 1992;67:1015–1030. [PubMed: 1597695]

Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. Behav Brain Res 2005;156:95–103. [PubMed: 15474654]

Schoups A, Vogels R, Qian N, Orban G. Practising orientation identification improves orientation coding in V1 neurons. Nature 2001;412:549–553. [PubMed: 11484056]

Seitz AR, Watanabe T. Psychophysics: Is subliminal learning really passive. Nature 2003;422:36. [PubMed: 12621425]

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci 2003;23:5545–5552. [PubMed: 12843255]

Shahin AJ, Roberts LE, Pantev C, Aziz M, Picton TW. Enhanced anterior-temporal processing for complex tones in musicians. Clin Neurophysiol 2007;118:209–220. [PubMed: 17095291]

Smith JC, Marsh JT, Brown WS. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr Clin Neurophysiol 1975;39:465–472. [PubMed: 52439]

Suga N, Gao E, Zhang Y, Ma X, Olsen JF. The corticofugal system for hearing: recent progress. Proc Natl Acad Sci U S A 2000;97:11807–11814. [PubMed: 11050213]

Tervaniemi M, Jacobsen T, Rottger S, Kujala T, Widmann A, Vainio M, Naatanen R, Schroger E. Selective tuning of cortical sound-feature processing by language experience. Eur J Neurosci 2006;23:2538–2541. [PubMed: 16706861]

Tremblay K, Kraus N, Carrell TD, McGee T. Central auditory system plasticity: generalization to novel stimuli following listening training. J Acoust Soc Am 1997;102:3762–3773. [PubMed: 9407668]

Tremblay K, Kraus N, McGee T, Ponton C, Otis B. Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. Ear Hear 2001;22:79–90. [PubMed: 11324846]

Trice R, Carrell T. Level 16. 1998

Watanabe T, Nanez JE, Sasaki Y. Perceptual learning without perception. Nature 2001;413:844–848. [PubMed: 11677607]

Wechsler, D. Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: The Psychological Corporation; 1999.

Wong PCM, Perrachione TK. Learning pitch patterns in lexical identification by native English-speaking adults. Applied Psycholinguistics. 2007

Wong PCM, Perrachione TK, Parrish TB. Neural characteristics of successful and less successful speech and word learning in adults. Human Brain Mapping. 2007a

Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci 2007b;10:420–422. [PubMed: 17351633]

Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. Science 1992;256:846–849. [PubMed: 1589767]

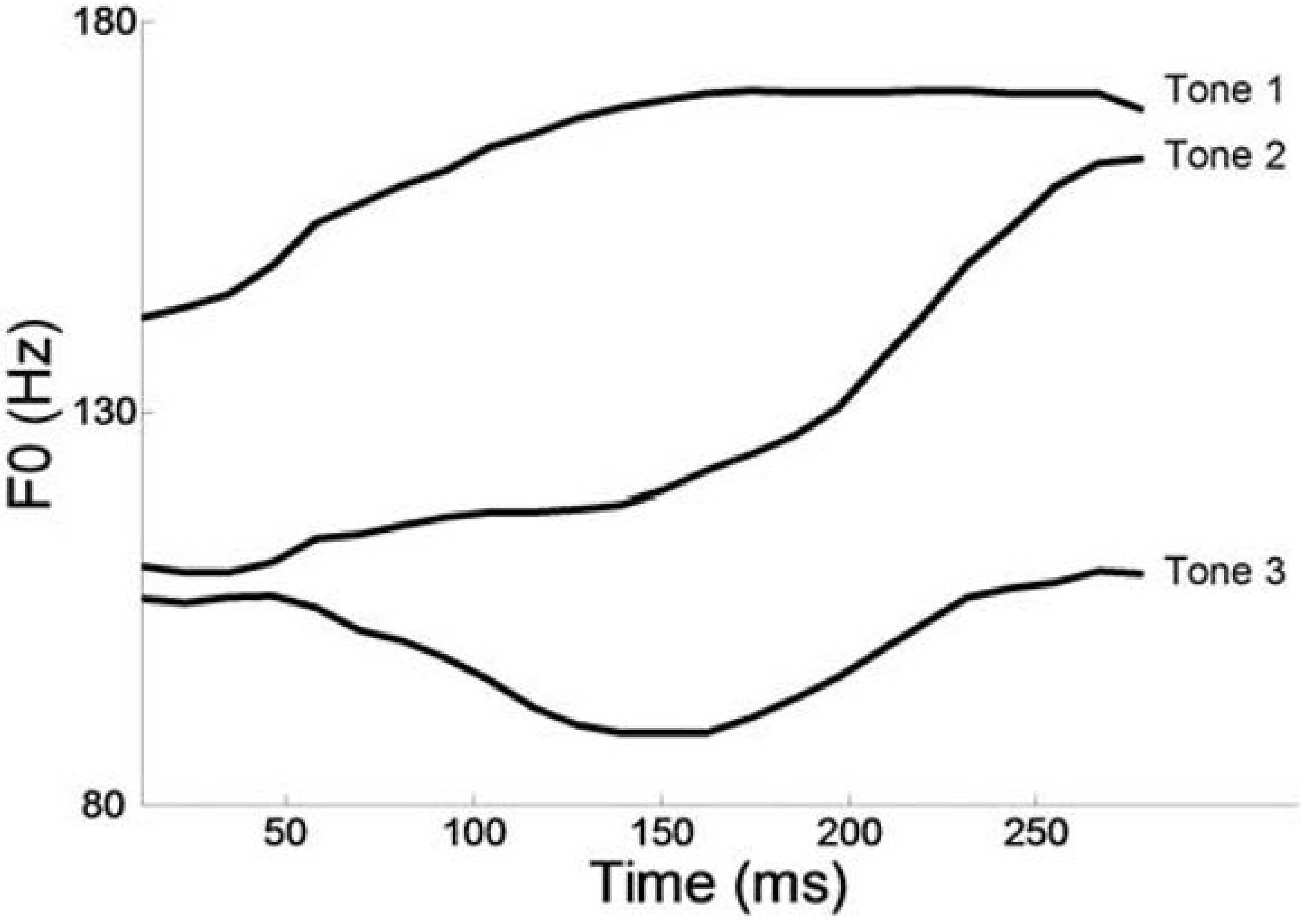

Fig. 1.
F0 contours of the high-level (Tone 1), rising (Tone 2), and dipping/falling-rising (Tone 3) patterns extracted from the /mi/ stimuli used for physiologic recording (F0 ranges: 140-172 Hz, 110-163 Hz, and 89-110 Hz, respectively).

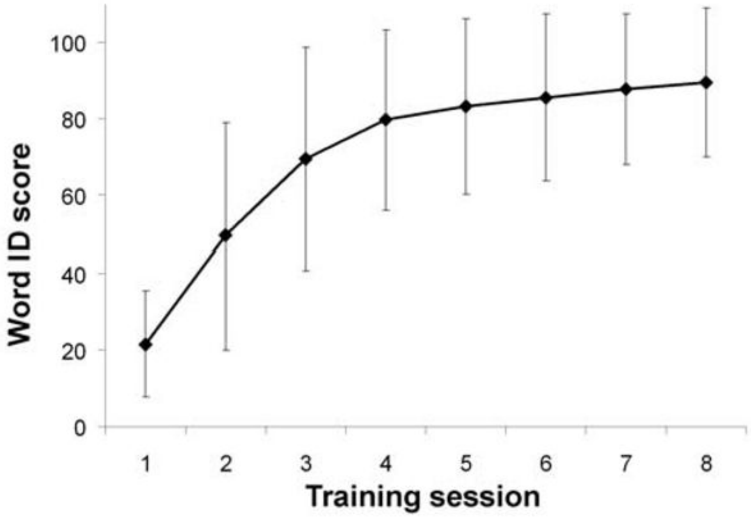

Fig. 2.
Mean word identification scores. Subjects' mean word identification scores measured at the end of each day of training indicating learning progress. Error bars show one standard deviation from the mean.

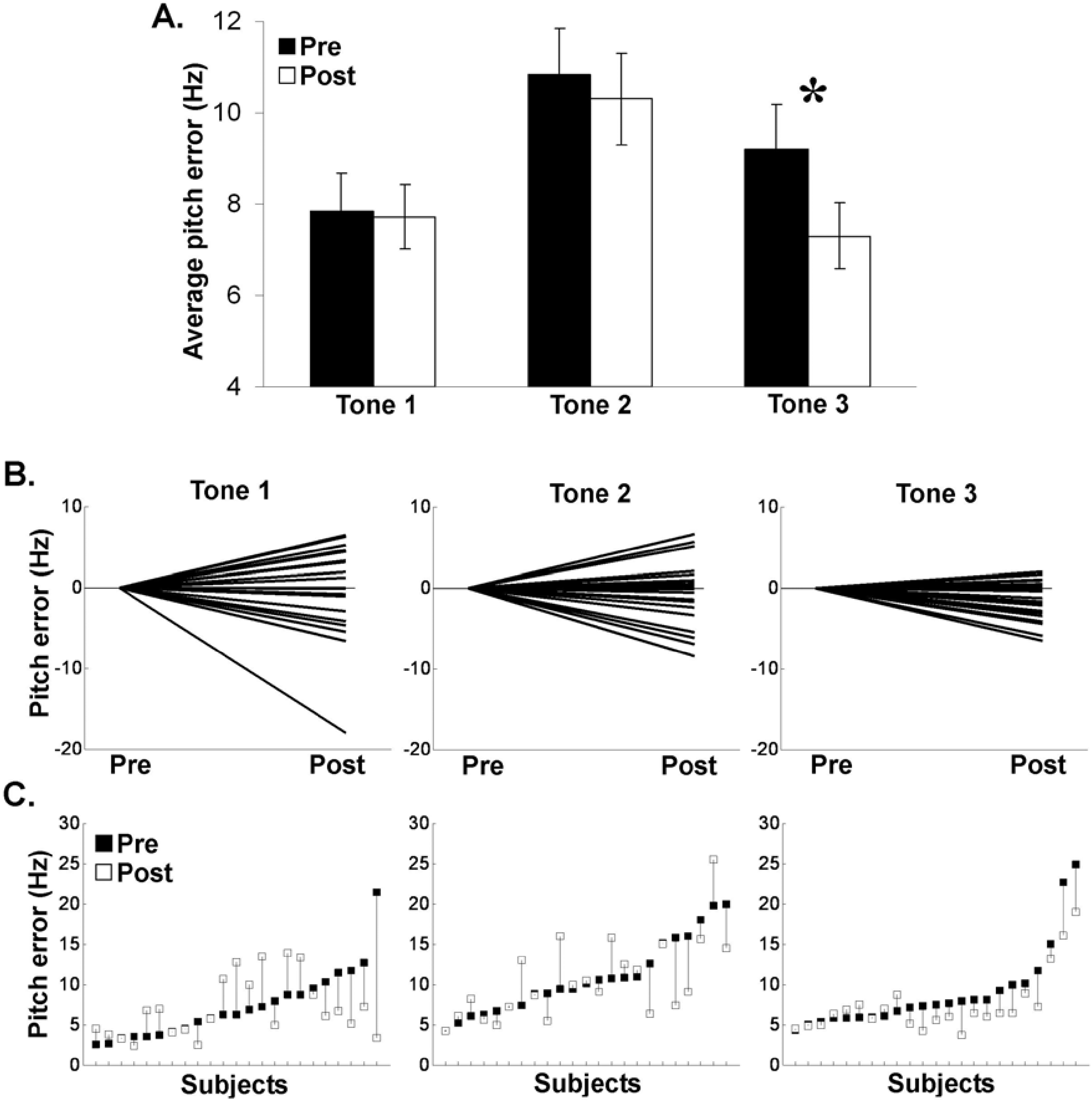

Fig. 3.

Pre- and post-training *Linear Pitch Tracking Error* for Tones 1, 2 and 3. (A) Average pre- and post-training *Linear Pitch Tracking Error* for each tone (± 1 SD). Note that the higher the bar, the larger the deviation from the stimulus contour in Hertz (Hz). *p < .0001 (B) Distribution of post- minus pre-training *Linear Pitch Tracking Error* for individual subjects (pre-training is plotted at zero, post-training is plotted as the difference between the pre- and post-training *Linear Pitch Tracking Error* values). In comparison to Tone 3, Tone 1 and Tone 2 show greater variability. (C) Dot plots of the individual *Linear Pitch Tracking Error* values before (black squares) and after training (white squares) with a vertical line connecting *Linear Pitch Tracking Error* values pre- and post-training for each subject.

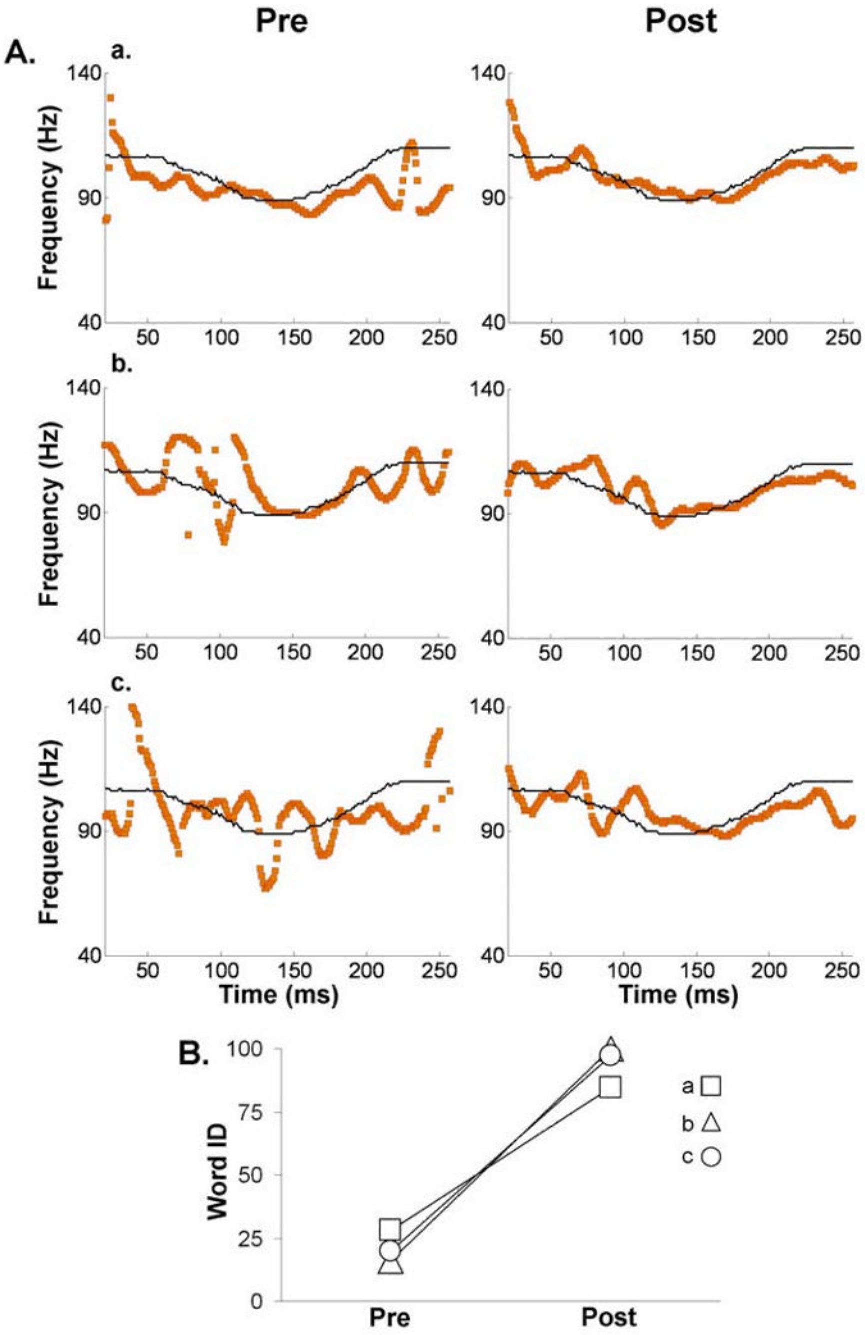

Fig. 4.

Representative examples of pre- and post-training pitch-tracking plots. (A) Trajectories (orange line) of brainstem pitch tracking elicited by a dipping pitch contour (Tone 3) for three subjects before and after training. The black line indicates the stimulus (expected) F0 contour. The time indicated on the *x*-axis refers to the midpoint of each 40-ms time bin analyzed. (B) Word identification scores of the first and last training sessions for the subjects plotted in panel A.

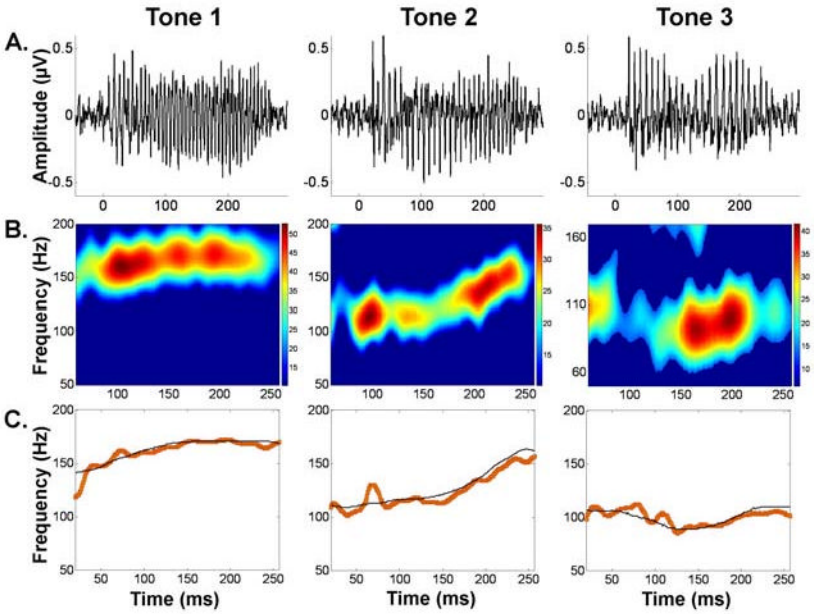

Fig. 5.

FFR waveforms, spectrograms, and pitch tracking plots for Tones 1, 2, and 3 for a representative subject. (A) FFR waveforms, (B) FFR spectra, and (C) pitch tracking plots with stimulus F0 contour (black) and response F0 contour (red) for Tones 1, 2 and 3 after training for a representative subject as a function of time referring to the midpoint of each 40-ms time bin analyzed. In the FFR spectra, color represents FFR spectral amplitude (arbitrary units). The stimulus F0 contours of the high-level (Tone 1), rising (Tone 2) and dipping/falling-rising (Tone 3) range between 140-172 Hz, 110-163 Hz, and 89-110 Hz, respectively. This subject's pre- and post-training *Linear Pitch Tracking Error* values are 2.58 and 4.77 (Tone 1), 14.74 and 6.61 (Tone 2), and 8.07 and 4.27 (Tone 3); S*pectral Dominance* values are 0 and 4 (Tone 1), 50 and 0 (Tone 2) and 2 and 0 (Tone 3), and *Pitch Noise Ratio* is 0 for both pre and post responses for all three tones. This subject is plotted in Fig. 4 as Subject b.

Table I

Training Stimuli. Subjects were trained on a vocabulary of 18 pseudowords. Each word, written in the International Phonetic Alphabet, is followed by its corresponding meaning in quotes. Numbers following lexical items designate tone. High-level tone is indicated by 1, rising tone by 2, and dipping tone by 3, according to convention (adapted from Wong and Perrachione, 2007).

p^{h}ɛŝ1 ‘glass’	dri1 ‘arm’	nɛr1 ‘boat’	vɛs1 ‘hat’	nΛk1 ‘brush’	fjut1 ‘shoe’
p^{h}ɛŝ2 ‘pencil’	dri2 ‘phone’	nɛr2 ‘potato’	vɛs2 ‘tape’	nΛk2 ‘tissue’	fjut2 ‘book’
p^{h}ɛŝ3 ‘table’	dri3 ‘cow’	nɛr3 ‘dog’	vɛs3 ‘piano’	nΛk3 ‘bus’	fjut3 ‘knife’

[Dear reviewers and editor: the online system does not convert IPA symbols probably. Here we use /ŝ/ to mean the voiceless postalveolar fricative as in ‘shirt’, and /r/ to mean the retroflex appoximant as in ‘row’ in American English]