Training to Improve Hearing Speech in Noise: Biological Mechanisms

Judy H. Song[1,2], Erika Skoe[1,2], Karen Banai[3] and Nina Kraus[1,2,4,5]

[1]Auditory Neuroscience Laboratory and [2]Department of Communication Sciences, Northwestern University, Evanston, IL 60208, USA, [3]Department of Communication Sciences and Disorders, University of Haifa, Haifa 31905, Israel, [4]Department of Neurobiology and Physiology and [5]Department of Otolaryngology, Northwestern University, Evanston, IL 60208, USA

Address correspondence to Dr Nina Kraus, Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA. Email: nkraus@northwestern.edu. http://www.brainvolts.northwestern.edu.

We investigated training-related improvements in listening in noise and the biological mechanisms mediating these improvements. Training-related malleability was examined using a program that incorporates cognitively based listening exercises to improve speech-in-noise perception. Before and after training, auditory brainstem responses to a speech syllable were recorded in quiet and multitalker noise from adults who ranged in their speech-in-noise perceptual ability. Controls did not undergo training but were tested at intervals equivalent to the trained subjects. Trained subjects exhibited significant improvements in speech-in-noise perception that were retained 6 months later. Subcortical responses in noise demonstrated training-related enhancements in the encoding of pitch-related cues (the fundamental frequency and the second harmonic), particularly for the time-varying portion of the syllable that is most vulnerable to perceptual disruption (the formant transition region). Subjects with the largest strength of pitch encoding at pretest showed the greatest perceptual improvement. Controls exhibited neither neurophysiological nor perceptual changes. We provide the first demonstration that short-term training can improve the neural representation of cues important for speech-in-noise perception. These results implicate and delineate biological mechanisms contributing to learning success, and they provide a conceptual advance to our understanding of the kind of training experiences that can influence sensory processing in adulthood.

Keywords: auditory training, brainstem encoding, fundamental frequency, LACE, Listening and Communication Enhancement, pitch encoding, speech-in-noise perception, speech perception

Introduction

In everyday listening situations, accurate speech perception relies on the capacity of the auditory system to process complex sounds in the presence of background noise. This process is often challenging even for young adults with normal hearing and normal cognitive abilities (Neff and Green 1987; Assmann and Summerfield 2004). In this study, we examined the extent to which speech-in-noise perception improved for young adult listeners who have undergone auditory training exercises that mimic real-world listening conditions, such as listening to a target speaker amid multiple background speakers. In addition, we measured the effects of this training on aspects of auditory brainstem processing that underlie speech perception in noise and the extent to which brainstem encoding of speech could predict learning success.

There is ample evidence of auditory training resulting in perceptual enhancements (Wright et al. 1997; Amitay et al. 2005; Moore et al. 2005; Mossbridge et al. 2006; Johnston et al.

2009) as well as plasticity in single neurons (Kraus and Disterhoft 1982; Diamond and Weinberger 1984, 1986, 1989) and neuronal populations (Olds et al. 1972; Bakin and Weinberger 1990; Recanzone et al. 1992; Edeline et al. 1993; Weinberger 1993; Gaab et al. 2006). In humans, learning-related cortical plasticity has been found after discrimination training using tones (Naatanen et al. 1993) and synthetic speech stimuli (Kraus et al. 1995; Tremblay et al. 2001, 2009). However, there have been surprisingly few investigations of how training impacts speech-in-noise perception (Burk and Humes 2007; Cainer et al. 2008; Yund and Woods 2010). These studies, which used small stimulus sets, indicate that while speech-in-noise perception can improve when training on words or sentences in artificial listening conditions, generalization to untrained materials is limited. These findings suggest that learning resulting from such training paradigms is specific to the trained speech materials and the parameters of the background noise (i.e., signal-to-noise ratio [SNR], babble vs. white noise, etc.) (Burk et al. 2006; Yund and Woods 2010). Nevertheless, long term auditory training (musical training) can improve speech perception under challenging conditions (Parbery-Clark et al. 2009a, Parbery-Clark et al. 2011, Bidelman and Krishnan 2010). To investigate the biological mechanisms driving neural changes under more naturalistic conditions, we used a training program that has been shown to benefit speech-in-noise perception in older adults. This commercially available program, "Listening and Communication Enhancement" (LACE), utilizes a large stimulus set (i.e., open-set speech material presented in a variety of difficult listening conditions often encountered in real life), incorporates feedback, and activates higher level cognitive skills—all of which have been suggested to promote perceptual learning and engender generalization (Kujala et al. 2001; Schaffler et al. 2004; Moore et al. 2005; Moore and Amitay 2007; Smith et al. 2009).

This is the first study to examine the effects of speech-in-noise training on human subcortical processing of sound. The auditory brainstem response (ABR), a noninvasive objective measurement of brainstem integrity (Hall 1992; Hood 1998), is ideal for examining the biological mechanisms underlying improvements in hearing in noise because the response is stable from session to session even when recorded in the presence of an acoustic masker (Russo et al. 2005; Song, Nicol, et al. 2010a). This response, which is reflective of synchronized potentials produced by populations of neurons along the subcortical auditory pathway (Møller and Jannetta 1985; Chandrasekaran and Kraus 2010), can be elicited by a wide range of acoustic stimuli, including speech syllables (King et al. 2002; Krishnan 2002; Galbraith et al. 2004; Krishnan et al. 2004, 2005; Russo et al. 2004; Xu et al. 2006; Wong, Skoe, et al. 2007; Aiken and Picton 2008; Akhoun et al. 2008; Swaminathan et al.

2008). Brainstem synchronization is so precise that disparities on the order of tenths of milliseconds are clinically significant (Jacobson 1985; Hood 1998). Consequently, even subtle training-related differences in the speech-evoked ABR could be indicative of meaningful neural plasticity.

To examine the malleability of the biological mechanisms underlying improvements in speech-in-noise perception, we measured ABRs from a group of young adults immediately before and after auditory training. Because long-term retention of behavioral and neurophysiologic changes has not been demonstrated previously, we also obtained neural and behavioral measurements 6 months after training. Behavioral and neurophysiological measures were also obtained from untrained control subjects at intervals equivalent to the trained group. By including nonnative English speakers into our subject pool, we obtained a wide range of performance on speech-in-noise tasks without confounding factors such as cognitive impairment or hearing loss. Nonnative English speakers tend to have more difficulty with English speech recognition in suboptimal acoustic environments compared with native English speakers even though their speech recognition ability is comparable in quiet listening conditions (Nabelek and Donahue 1984; Takata and Nabelek 1990; Mayo et al. 1997; Meador et al. 2000; Rogers et al. 2006).

We hypothesized that auditory training would yield improvements in speech-in-noise perception by strengthening brainstem encoding of acoustic features that are crucial for speech-in-noise perception. Specifically, we predicted that the encoding of the fundamental frequency (F_0) and the second harmonic (H_2) would be enhanced with training. This is because successful tracking of the target message is reliant on the listener's ability to benefit from these pitch cues found in speech (Brokx and Nooteboom 1982; Assmann and Summerfield 1990; Meddis and O'Mard 1997; Bird and Darwin 1998). This hypothesis is supported by recent work highlighting the relationship between speech-in-noise perception and subcortical representation of pitch cues (Anderson et al. 2010; Anderson et al. 2011; Song, Skoe, et al. 2010b). It is also known that subcortical experience-dependent plasticity occurs over the lifetime (Krishnan et al. 2005; Musacchia et al. 2007; Wong, Skoe, et al. 2007; Parbery-Clark et al. 2009b; Kraus and Chandrasekaran 2010) and following short-term auditory training (Russo et al. 2005; de Boer and Thornton 2008; Song et al. 2008; Carcagno and Plack 2010; Kumar et al. 2010), although subcortical plasticity following speech-in-noise training has not been previously evaluated.

Materials and Methods

Participants

Sixty young adults (38 females), aged 19–35 years (mean age = 24.7, standard deviation [SD] = 3.5 years) with no history of neurological disorders, participated in this study. To control for musicianship, a factor known to modulate frequency encoding at the level of the brainstem (Parbery-Clark et al. 2009b; Strait et al. 2009; Kraus and Chandrasekaran 2010), all subjects had fewer than 6 years of musical training that ceased 10 or more years prior to study enrollment. All participants had normal IQ (mean SD: 106, SD = 9) as measured by the Test of Nonverbal Intelligence-3 (Brown et al. 1997), normal hearing (≤20 dB Hearing Level [HL] pure-tone thresholds from 125–8000 Hz), and normal click-evoked ABR wave V latencies to 100 μs clicks presented at 31.1 Hz at 80.3 dB sound pressure level (SPL). Participants gave their informed consent in accordance with the Northwestern University Institutional Review Board.

Participants were randomly assigned to a group that underwent training or to a group that received no remediation. The trained group was composed of 28 young adults (17 females), aged 19–35 years (mean age = 26.0, SD = 3.8 years), and the control group was composed of 32 young adults (21 females), aged 20–31 years (mean age = 23.7, SD = 2.8 years). For both the trained and control groups, half of the participants were nonnative English speakers, with the other half consisting of native English speakers. Nonnative speakers were included to widen the range of performance on the English-based speech-in-noise tasks for this normal hearing population. Nonnative trained and nonnative control groups were matched on English proficiency. All nonnative speakers started learning English after 5 years of age, resided in the United States of America, and had good English proficiency (score ≥ 50) as measured by the Speaking Proficiency English Assessment Kit score (mean score = 52.61, SD = 7.4) and the Peabody Picture Vocabulary Test score (mean score = 107.92, SD = 31.2).

For the trained group, the mean test time between pre- and posttraining sessions was 57.43 (±34.01) days. The control group was tested at time intervals equivalent to the trained participants with the mean time between Test 1 and Test 2 being 57.5 (±32.8) days ($t = 0.001$, $P = 0.99$). Approximately 6 months after the completion of LACE (mean time: 6 months ± 3 weeks), a subset of the original subjects (trained group $n = 16$, control group $n = 12$) returned to the laboratory and were retested a third time (Test 3) to examine the retention of training-related changes.

Neurophysiologic Stimuli and Recording Parameters

Brainstem responses were elicited in response to the syllable [da] in a quiet and a multitalker noise condition. This stimulus was a six-formant syllable synthesized at a 20 kHz sampling rate using a Klatt synthesizer (Klatt 1980). The duration was 170 ms with voicing ($F_0 = 100$ Hz) onset at 10 ms (Fig. 1A). The formant transition duration was 50 ms and comprised a linearly rising first formant (F_1) (400–720 Hz), linearly falling F_2 and F_3 (1700–1240 Hz and 2580–2500 Hz, respectively), and flat F_4 (3300 Hz), F_5 (3750 Hz), and F_6 (4900 Hz). After the transition region, these formant frequencies remained constant at 720, 1240, 2500, 3300, 3750, and 4900 Hz for the remainder of the syllable (50–170 ms). The stop burst consisted of 10 ms of initial frication centered at frequencies around F_4 and F_5.
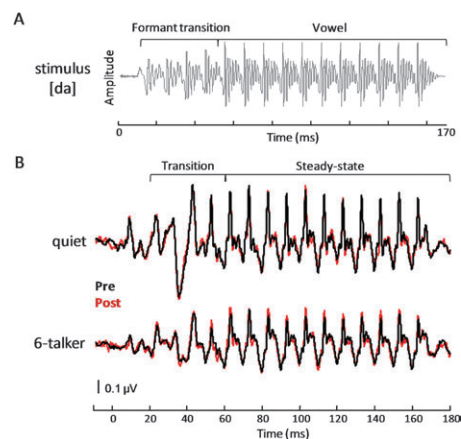

Figure 1. (A) The acoustic waveform of the target stimulus [da]. The formant transition and vowel regions are bracketed. The periodic amplitude modulations of the stimulus, reflecting the rate of the fundamental frequency ($F_0 = 100$ Hz), are represented by the major peaks in the stimulus waveform (10 ms apart). (B) Pre- (black) and posttraining (red) grand average brainstem responses of trained subjects recorded to the [da] stimulus in quiet (top) and in the 6-talker babble (bottom) condition over the entire response. Overlay of pre- and posttraining responses shows that the posttraining brainstem response recorded in 6-talker babble exhibits a more robust representation of F_0. This enhancement is demonstrated by larger amplitudes of the prominent periodic peaks occurring every 10 ms in this time region. The transition portion of the response reflects the shift in formants of the stimulus as it moves from the onset burst of the stop consonant to the vowel portion of the syllable. The steady-state portion of the response reflects phase locking to stimulus periodicity in the vowel.

The syllable [da] was presented at 80.3 dB SPL at a rate of 4.35 Hz in alternating polarities via an insert earphone placed in the right ear (ER-3; Etymotic Research, Elk Grove Village, IL).

The noise condition consisted of 6-talker (3 female and 3 male native English speakers) babble spoken in English. To create the babble, the speakers were instructed to speak in a natural conversational style. Recordings of nonsense sentences were made in a sound-attenuated booth in the phonetics laboratory of the Department of Linguistics at Northwestern University (Smiljanic and Bradlow 2005) and were digitized at a sampling rate of 16 kHz with a 24-bit accuracy (for further details, see Smiljanic and Bradlow 2005; Van Engen and Bradlow 2007). The tracks were root mean square amplitude normalized using Level 16 software (Tice and Carrell 1998). The 6-talker babble track (4000 ms) was looped for the duration of data collection (ca. 25 min) with no silent intervals. This presentation paradigm allowed the noise to occur at a randomized phase with respect to the target speech sound [da]. Thus, responses that were time locked to the target sound could be averaged without the confound of including phase coherent responses to the background noise.

In the quiet and noise conditions, 6300 sweeps of the ABR to [da] were collected using Scan 4.3 Acquire (Compumedics, Charlotte, NC) in continuous mode at a sampling rate of 20 kHz. The continuous recordings were filtered, artifact (±35 μV) rejected, and averaged off-line using Scan 4.3. Responses were band-pass filtered from 70 to 1000 Hz (12 dB/octave) to isolate brainstem activity. Waveforms were averaged with a time window spanning 40 ms prior to the onset and 16.5 ms after the offset of the stimulus and then baseline corrected over the prestimulus interval (–40 to 0 ms, with 0 corresponding to the stimulus onset). Responses of alternating polarity were added to isolate the neural response by minimizing stimulus artifact and the cochlear microphonic (Gorga et al. 1985). The final average response consisted of the first 6000 artifact-free responses.

All responses were differentially recorded using Ag–AgCl electrodes from Cz (active) to right earlobe (reference), with forehead as ground. During testing, the participants watched a subtitled video of their choice to facilitate a passive yet wakeful state.

Neurophysiologic Analysis Procedures
In order to examine the impact of auditory training on the subcortical representation of pitch information, we analyzed the frequency following response (FFR), a component of the brainstem response that reflects the encoding of periodic elements in the acoustic stimuli (Moushegian et al. 1973; Chandrasekaran and Kraus 2010). The FFR reflects specific spectral and temporal properties of the signal with such fidelity that interpeak intervals are synchronized to the period of the fundamental frequency and its harmonics (Galbraith et al. 1995). This enables comparisons between the frequency composition of the response and the corresponding features of the stimulus (Galbraith et al. 2000; Russo et al. 2004; Kraus and Nicol 2005; Skoe and Kraus 2010). Thus, the FFR offers an objective measure of the degree to which the auditory pathway accurately encodes complex acoustic features, including those known to play a role in characterizing speech under adverse listening conditions. Moreover, the FFR is sensitive to the masking effects of competing sounds (Yamada et al. 1979; Ananthanarayan and Durrant 1992; Russo et al. 2004, 2009; Wilson and Krishnan 2005; Parbery-Clark et al. 2009b; Anderson et al. 2010; Krishnan et al. 2010; Li and Jeng 2011), resulting in delayed and diminished responses (i.e., neural asynchrony). These features of the response make it well suited for examining the impact that a training program may have on the encoding of speech cues in noise, particularly those that convey pitch information.

Regions of Interest (20–60 and 60–180 ms)
The FFR was divided into 2 regions for analysis: 1) A transition region (20–60 ms) corresponding to the encoding of the time-varying transition of the stimulus as the syllable [da] proceeds from the stop consonant to the vowel and 2) a steady-state region (60–180 ms) corresponding to the encoding of the vowel (Fig. 1*B*). These regions were analyzed separately because 1) temporal cues relating to the F_0 are more variable during the formant transition compared with the steady-state region (Song, Skoe, et al. 2010b), 2) recent studies have shown that speech-in-noise perception correlates with subcortical

encoding of pitch cues during the formant transition but not the steady-state region (Anderson et al. 2010; Song, Skoe, et al. 2010b), and 3) rapidly changing formant transitions pose particular perceptual challenges (Miller and Nicely 1955; Tallal and Piercy 1974; Merzenich et al. 1996; Hedrick and Younger 2007). Since the time-varying formant transition of stop consonants is known to be perceptually vulnerable in background noise, we expected training-associated improvements in the neural representation of the F_0 and H_2 might be most evident in the response to the formant transition.

Fast Fourier Analysis
The spectral energy in the frequency domain was analyzed in MATLAB version 2007b (The Mathworks, Inc., Natick, MA) by computing fast Fourier transforms with zero padding to increase the resolution of the spectral display (Skoe and Kraus 2010). Average spectral amplitudes were calculated for the formant transition region and steady-state region using 40 Hz bins centered around the F_0 (100 Hz) and harmonics (integer multiples of 100 Hz). Two composite scores were generated: 1) F_0 and H_2 amplitudes were averaged to obtain the strength of pitch cue encoding and 2) the amplitudes of the 3rd–10th harmonics (H_3–H_{10}: 300–1000 Hz) were averaged to generate a measure of overall harmonic encoding.

Analysis of Experience-Dependent Plasticity
Training-related differences were measured using a repeated measures analysis of variance with test session as the within-subject factor and group as a between-subject factor. Post hoc pairwise tests were performed to establish in which group the significant changes occurred. For a given neurophysiological measure, changes in the trained group that significantly exceeded those in the untrained control group were considered evidence of training-related neuroplasticity.

Auditory Training
The LACE program (Neurotone, Inc., 2005) is an adaptive computer-based program that consists of 20 training sessions, approximately 30 min each in duration. Training was completed over the course of 4 weeks, and participants were required to complete 5 sessions a week. LACE provided participants with practice on tasks falling into 3 categories: 1) comprehension of degraded speech, 2) cognitive skills, and 3) communication strategies. Degraded speech tasks included speech in multitalker noise (Sentences-in-Noise), speech with one competing speaker (Competing Speaker), and time-compressed speech (Rapid Speech). Cognitive skill tasks included missing word (Auditory Closure) and target word exercises (Auditory Memory). Each task was given at interspersed intervals throughout the 20 training sessions such that no one session included all tasks. At the conclusion of each exercise, participants were presented with "helpful hints," communication strategies for a variety of situations including adapting the acoustical environment for optimal communication and using assertive listening and communication techniques. LACE provided immediate visual feedback during each exercise and at the completion of each training session. Composite scores were calculated based on the performance on the LACE subcomponents (i.e., Competing Speaker, Rapid Speech, Auditory Memory, and Auditory Closure) and were displayed at the end of each session, enabling the trainees to track their overall progress.

All the trained participants installed the LACE program from a CD-ROM onto their personal computers and underwent the training at home. Stimuli were presented in sound field via 2 speakers placed on either side of the computer and facing the participant. The sound level was set at a comfortable listening level, as determined by the trainee, prior to the start of each training session. At the end of each session, results were uploaded to a centralized website. Compliance with the training regimen was monitored by the experimenter (first author) through this website.

Perceptual Abilities Testing (Speech-in-Noise)
Quick Speech-in-Noise Test (QuickSIN; Etymotic Research) (Killion et al. 2004) is a nonadaptive test of speech perception in multitalker noise. During QuickSIN testing, each participant was given one practice

list to acquaint him/her with the task. Then 4 of the 18 lists of the QuickSIN were randomly selected and administered. Each list consisted of 6 target sentences spoken by a female talker that were embedded in 4-talker babble (3 women and 1 man). The sentences were grammatically correct yet did not contain many semantic or contextual cues (Wilson et al. 2007). Sample sentences, with key words italicized, included, "A *force equal* to that *would move* the *earth*." and "The *weight* of the *package* was *seen* on a *high scale*." The SNR decreased in 5 dB increments from 25 to 0 dB SNR with each sentence presentation. This test sentences were presented binaurally at 70 dB HL via ER-2 insert earphones (Etymotic Research). Instructions and feedback were also delivered through the insert earphones. The QuickSIN CD was played through a compact disc player connected externally to a GSI 61 audiometer. Participants were instructed to repeat back sentences spoken by the target talker by speaking into a microphone that was mounted on a wall directly in front of them. The tester also adjusted the intensity of the subjects' voice to a comfortable listening level using the "Talkback" dial on the audiometer. In scoring the test, one point was awarded for each of the 5 key words that were repeated correctly for each sentence. Incorrect or omitted words were annotated on a response form by the tester, and performance scores were checked by an independent auditor blind to subject grouping. The SNR loss score for each list was calculated following the instruction manual for the QuickSIN test by adding the number of key words repeated correctly, summing across all 6 sentences, and subtracting the total correct from 25.5. A more negative SNR loss was indicative of better performance on the task. The 4 SNR loss scores were averaged to obtain each participant's final QuickSIN SNR loss score for each test session.

The Hearing in Noise Test (HINT; Biologic Systems Corp., Mundelein, IL) (Nilsson et al. 1994) is an adaptive test that measures speech perception ability. During the HINT, participants were instructed to repeat back short, semantically, and syntactically simple sentences by speaking into a microphone that they held a few inches from their mouths. The HINT sentences were presented in a speech-shaped background noise that matched the spectra of the test sentences. No background noise was present at the time the participants provided their responses orally. Sample sentences included "Sugar is very sweet" and "Children like strawberries." The speech stimuli consisted of Bamford–Kowal–Bench (1979) sentences (12 lists of 20 sentences) spoken by a man and presented in sound field via 2 loudspeakers positioned at a 90° angle from each other. The participant sat 1 m equidistant from the 2 loudspeakers, and the target sentences were always presented from the speaker that he/she was facing (source location of 0°) for each listening condition. There were 3 noise conditions that only differed in the location of the speaker presenting the noise: noise delivered from 0° (HINT-Front), 90° to the right (HINT-Right), and 90° to the left (HINT-Left). The intensity level of the noise was fixed at 65 dB SPL, and the intensity level of the target sentences was adjusted by the HINT software until a threshold SNR was obtained (i.e., difference in intensity [dB] between the speech and noise level for which 50% of sentences are correctly repeated). Each participant was given 4 practice sentences in the HINT-Front condition at the beginning of the test before they were administered each of the 3 noise conditions. The tester adjusted the intensity of the subjects' voice to a comfortable listening level using the computer's volume setting. Threshold SNRs were calculated for the 3 conditions with a more negative SNR indicating a greater ability to perceive speech in more adverse listening conditions. Composite scores were also calculated based on the threshold SNRs of the 3 HINT noise conditions. During the pre- and posttesting sessions, participants were randomly presented with a different sentence list for each listening condition.

Neural Correlates of Learning

We examined relationships between training-related brainstem plasticity and changes in perceptual measures and potential neural indicators of behavioral improvement. To examine the relationship between subcortical encoding of pitch- and training-associated learning, composite measures obtained from the average of F_0–H_2 amplitudes, derived from the transition and steady-state response regions were correlated with the amount of learning on LACE tasks and on independent measures of speech-in-noise perception (e.g., QuickSIN SNR loss [post]–QuickSIN SNR loss [pre]).

Results

General Auditory Training

To determine if training-related changes in LACE performance were statistically significant, initial performance scores were compared with scores obtained from the final session of each task. A repeated measures analysis of variance (RMANOVA) with training session (2: initial vs. final) and task (5: individual LACE components) as within-subject factors showed a significant main effect of task ($F_{4,24} = 227.707$, $P < 0.001$), as well as a significant interaction between training session and task ($F_{4,24} = 29.658$, $P < 0.001$). Post hoc paired t-test analysis showed a significant improvement in scores corresponding to Speech-in-Noise, Competing Speaker, Rapid Speech, Auditory Memory, and Auditory Closure tasks ($t_{27} = 4.981$, -9.550, 7.551, -10.814, -2.320, and $P = <0.001$, <0.001, <0.001, <0.001, 0.028, respectively). Moreover, since the Composite score of the first 2 sessions was always 90, the 'initial' Composite score used for analyses was obtained from the third training session and was compared with the final Composite score. A significant improvement on the Composite score, which reflects performance on all 5 listening tasks, was found ($t_{27} = -2.790$, $P = 0.010$).

Independent Measures of Speech-in-Noise Perception

On both independent measures of speech-in-noise perception, performance improved significantly between the first and the second test sessions in the trained group relative to the control group (Fig. 2). A 2 treatment group (Trained vs. Control) × 2 test session (Test 1 vs. 2) multivariate RMANOVA was performed on 4 dependent measures of speech-in-noise perception (i.e., QuickSIN loss scores and HINT Noise-Front, -Right, and -Left SNRs). Results showed a main effect of test session ($F_{4,55} = 10.074$, $P < 0.001$) and a significant interaction between test session and treatment group ($F_{4,55} = 7.065$, $P < 0.001$). Post hoc 2 treatment group (Trained vs. Control) × 2 test session (Test 1 vs. 2) univariate RMANOVA performed separately for QuickSIN loss scores, and HINT SNRs showed a main effect of test session (QuickSIN: $F_{1,58} = 27.299$, $P < 0.001$; HINT: $F_{1,58} = 16.760$, $P < 0.001$) and a significant interaction between test session and treatment group (QuickSIN: $F_{1,58} = 14.371$, $P < 0.001$, HINT: $F_{1,58} = 16.791$, $P < 0.001$). Follow-up paired t-test analyses for the trained group showed significant changes between the first and second sessions in the QuickSIN SNR loss scores ($t_{27} = 5.025$, $P < 0.001$) and in the HINT SNR values in the Noise-Front, -Right, and Composite conditions ($t_{27} = -3.586$, -4.280, -5.092; and $P = 0.001$, <0.001, <0.001, respectively). The HINT-Left condition showed a marginal change ($t_{27} = -1.926$, $P = 0.065$). Furthermore, the group that underwent training showed a significant relationship between initial performance on QuickSIN and subsequent amount of learning on LACE, such that better initial performance was related with less improvement ($r = -0.569$, $P = 0.002$) (Fig. 3).

For the control group, paired t-test analysis showed no significant changes between the first and second sessions in the
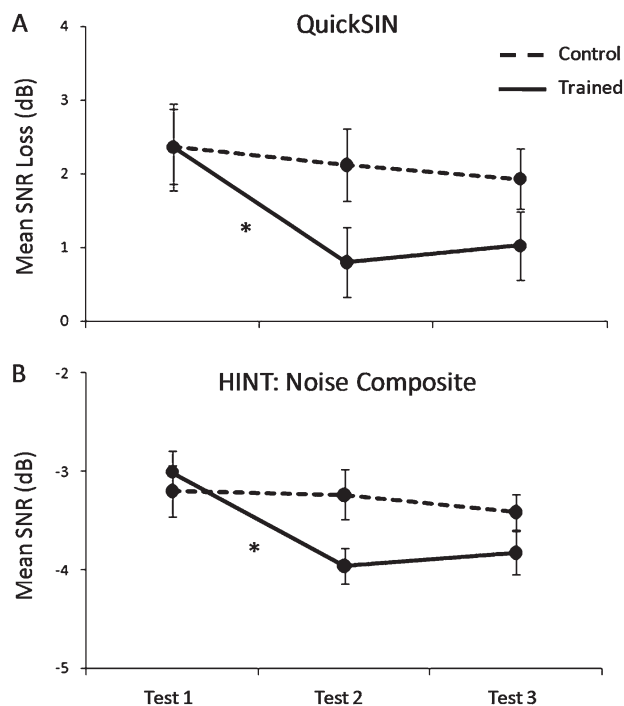
Figure 2. Mean speech-in-noise performance scores obtained from Tests 1, 2, and 3. Error bars represent one standard error of the mean (SEM). *$P < 0.05$. (*A*) Average QuickSIN SNR loss scores of the trained (solid) and control (dashed) groups from each test session. Trained subjects showed a significant improvement after training ($t_{27} = 5.025$, $P < 0.001$) on this task. (*B*) Average HINT Noise Composite SNR values of trained (solid) and control (dashed) groups from each test session. Trained subjects showed significant improvements after training in the Noise-Front, -Right, and Composite conditions ($t_{27} = -3.586$, -4.280, -5.092; and $P = 0.001$, <0.001, <0.001, respectively). The HINT-Left condition showed a marginal change ($t_{27} = -1.926$, $P = 0.065$).
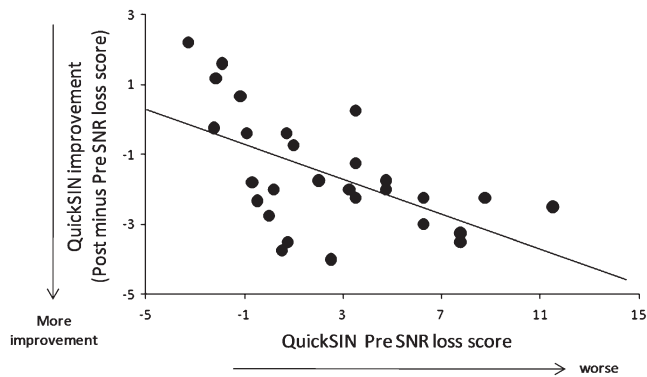

Figure 3. Amount of training-related improvement on the QuickSIN test plotted as a function of initial SNR loss threshold for each subject who trained on LACE ($r = -0.569$, $P = 0.002$).

QuickSIN SNR loss scores ($t_{31} = 1.407$, $P = 0.170$) nor in the HINT SNR values in the Noise-Front, -Right, -Left and Composite conditions ($t_{31} = -0.630$, 0.230, 0.290, -0.278; and $P = 0.533$, 0.820, 0.774, 0.783, respectively).

Retention of Speech-in-Noise Perceptual Improvements after 6 Months

The improvement in speech-in-noise perception was retained approximately 6 months after the initial retest session (Test 2).

A 2 treatment group (Trained vs. Control) × 3 test sessions (Tests 1, 2, and 3) multivariate RMANOVA performed on pre- and post-QuickSIN loss scores and HINT Noise-Front, -Right, and -Left SNRs as the dependent measures showed a main effect of test session ($F_{1,26} = 2.461$, $P = 0.018$) and approached a significant interaction between test session and treatment group ($F_{1,26} = 1.997$, $P = 0.054$).

Post hoc 2 treatment group (Trained vs. Control) × 2 test session (Test 2 vs. Test 3) univariate RMANOVA performed separately for QuickSIN loss scores and HINT SNRs did not show a significant main effect of test session (QuickSIN: $F_{1,26} = 1.951$, $P = 0.173$; HINT: $F_{1,26} = 0.002$, $P = 0.961$) nor a significant interaction between test session and treatment group (QuickSIN: $F_{1,26} = 0.027$, $P = 0.871$; HINT: $F_{1,26} = 0.379$, $P = 686$). This suggests that the perceptual improvement in the trained group was maintained several months after training was completed, while the control group demonstrated reliable test–retest speech-in-noise recognition ability across multiple test sessions.

Additionally, post hoc 2 treatment group (Trained vs. Control) × 2 test session (Test 1 vs. 3) univariate RMANOVA was performed separately for each speech-in-noise test. Similar to findings observed during Test 2, QuickSIN showed a significant a main effect of test session (QuickSIN: $F_{1,26} = 4.731$, $P = 0.038$) and a significant interaction between test session and treatment group (QuickSIN: $F_{1,26} = 5.516$, $P = 0.026$). The HINT showed a main effect of training that was approaching statistical significance ($F = 3.961$, $P = 0.057$) and a significant interaction between test session and treatment group ($F = 5.601$, $P = 0.010$). Furthermore, follow-up paired t-test analyses of speech-in-noise perceptual scores from test sessions 1 and 3 of the trained group showed significant improvement in the QuickSIN SNR loss scores ($t_{15} = 2.998$, $P = 0.008$) and in the HINT SNR values in the Noise-Front and Noise-Right conditions ($t_{15} = -2.694$, -2.800; and $P = 0.021$, 0.017; respectively). The HINT-Left condition continued to show a marginal change ($t_{15} = 0.533$, $P = 0.062$). The control group showed no significant change in performance (QuickSIN $t_{11} = -0.190$, $P = 0.853$, HINT-Front: $t_{11} = -0.799$, $P = 0.437$, HINT-Right: $t_{11} = -0.106$, $P = 0.917$, HINT-Left: $t_{11} = -1.089$, $P = 0.293$).

Neurophysiological Responses from the Brainstem

Stability of the Brainstem Measures Overtime

Test–retest data were collected from control subjects who did not undergo auditory training (refer to Song et al. (2010a) for more information on test–retest reliability of speech-evoked ABRs). RMANOVA revealed that the [da]-evoked brainstem responses recorded in quiet and background noise were stable and replicable between 2 sessions separated by approximately 2 months (Fig. 4).

Effects of Training Brainstem Measures

In the transition and steady-state regions of the response (20–60 and 60–180 ms) in background noise, we found significant training-related enhancements in pitch-related encoding after the auditory training program (Fig. 5). A 2 treatment group (Trained vs. Control) × 2 test session (Test 1 vs. 2) multivariate RMANOVA was performed on the composite measure of pitch (F_0 and H_2) encoding obtained from the transition and steady-state regions as dependent variables. A significant interaction was found between test session and treatment group for the
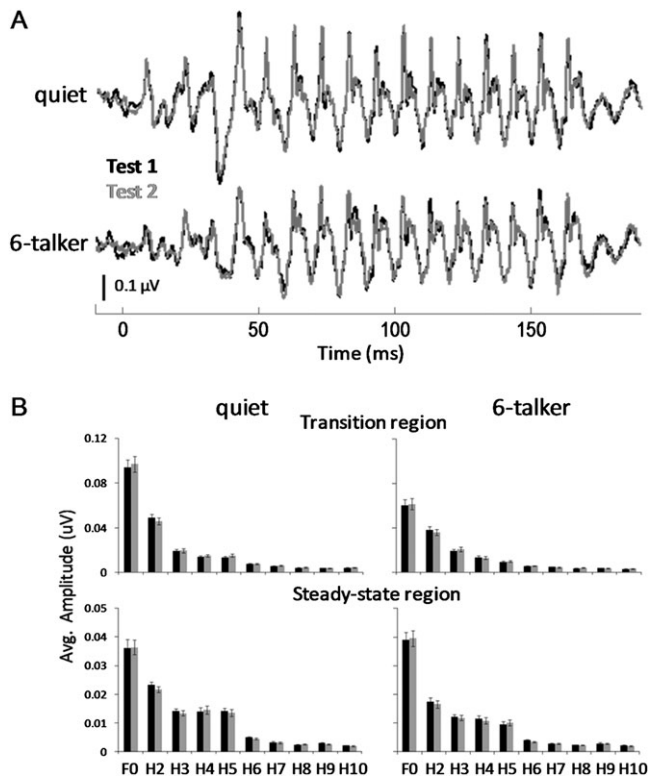
Figure 4. (*A*) Overlay of control groups' grand average brainstem responses from Test 1 (black) and Test 2 (gray) recorded in the quiet (top) and in 6-talker babble (bottom) conditions. (*B*) Grand average amplitudes of the fundamental frequency (100 Hz) and harmonics (200–1000 Hz) for the transition (top) and steady-state (bottom) regions of the control groups' response recorded in quiet and 6-talker babble obtained from Tests 1 (black) and 2 (gray) ($\pm$1 SEM).
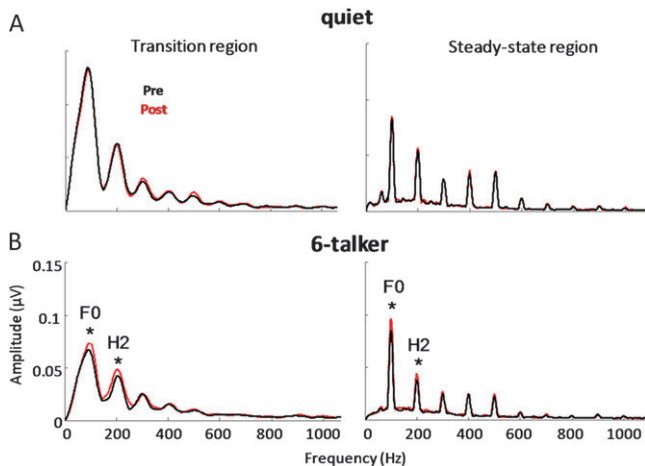

Figure 5. Grand average spectra of the F_0 (100 Hz) and harmonics (200–1000 Hz) calculated from responses recorded in (*A*) quiet and (*B*) 6-talker babble obtained from pre- (black) and posttraining (red) sessions ($\pm$1 SEM). Training-related enhancements in the representation of pitch cues (i.e., F_0 and H_2) were found in background noise in the trained group but not the control group as demonstrated by a significant interaction between test session and treatment group ($F_{1,58} = 6.498$, $P = 0.013$). This improvement was not seen in the brainstem response recorded in quiet (transition: $F_{1,58} = 0.034$, $P = 0.854$; steady-state: $F_{1,58} = 0.003$, $P = 0.960$).

pitch composite measure with the spectral magnitude for the F_0 and H_2 being enhanced after training in noise ($F_{1,58} = 6.498$, $P = 0.013$). Post hoc 2 treatment group (Trained vs. Control) ×

2 test session (Test 1 vs. 2) univariate RMANOVA was performed separately for the transition and steady-state regions to further assess pitch-related enhancement following training. Findings showed a significant interaction between test session and treatment group in each response region (transition region: $F_{1,58} = 6.498$, $P = 0.013$; steady-state region: $F_{1,58} = 9.174$, $P = 0.004$). Follow-up paired *t*-tests revealed a significant effect of training for F_0 (transition: $t_{27} = -2.103$, $P = 0.045$; steady-state: $t_{27} = -3.534$, $P = 0.001$) and H_2 ($t_{27} = -2.230$, $P = 0.034$; steady-state: $t_{27} = -6.614$, $P < 0.001$).

A 2 treatment group (Trained vs. Control) × 2 test session (Test 1 vs. 2) multivariate RMANOVA was performed on the composite measure of higher harmonic frequency (H_3–H_{10}) encoding obtained from the transition and steady-state regions as dependent variables. There was no main effect of training ($F_{1,58} = 0.731$, $P = 0.396$) nor significant interactions between test session and treatment group ($F_{1,58} = 0.333$, $P = 0.566$).

In the noise condition, training-related pitch cue enhancements were significantly different between the formant transition and steady-state regions (postminus preaverage F_0–H_2 magnitude) as shown by a significant 3-way interaction between the effect of training, response region, and frequency encoding ($F_{1,27} = 9.433$, $P = 0.009$). Greater enhancements were found for the transition region (amount of change: 0.00839 μV) of the response compared with the sustained vowel portion (amount of change: 0.00393 μV).

In the quiet condition, a 2 treatment group (Trained vs. Control) × 2 test session (Test 1 vs. 2) multivariate RMANOVA performed separately on the composite measures of pitch and higher harmonic frequency encoding obtained from both the transition and steady-state regions showed no training-related enhancements in the encoding of pitch cues ($F_{1,58} = -0.031$, $P = 0.861$) nor higher harmonics ($F_{1,58} = 0.426$, $P = 0.516$). There were also no significant interactions between test session and treatment group for either composite measures of frequency encoding (pitch cues: $F_{1,58} = 0.018$, $P = 0.896$; high harmonics: $F_{1,58} = 1.488$, $P = 0.227$).

Retention of Neurophysiologic Enhancement after 6 Months

The retention of training-related enhancements in the trained group was assessed using an RMANOVA comparing neural strength of spectral encoding (pitch and harmonics) across all 3 test sessions over the 2 response regions. This analysis showed main effects of training and region ($F_{1,27} = 24.351$, 40.960; $P < 0.001$, <0.001, respectively) as well as a significant interaction between training and time region ($F_{1,27} = 24.687$, $P < 0.001$). Post hoc paired *t*-tests revealed no significant change in the strength of neural encoding of pitch in either the transition (quiet: $F_{1,27} = 1.222$, $P = 0.329$; 6-talker babble: $F_{1,27} = 1.218$, $P = 0.330$) or the steady-state region (quiet: $F_{1,27} = 2.423$, $P = 0.131$; 6-talker babble: $F_{1,27} = 0.901$, $P = 0.430$) from retest sessions 2 and 3. Moreover, there was no significant change in the representation of higher harmonics (300–1000 Hz) in either the quiet or 6-talker babble conditions ($F_{1,27} = 1.267$, $P = 0.317$; $F_{1,27} = 0.479$, $P = 0.575$, respectively). These findings demonstrate long-term retention of the neurophysiologic enhancements seen immediately after completing training.

Relationship to Behavior

Participants with the largest strength of pitch encoding at pretest in the noise condition showed the greatest amount of

learning. Significant correlations were found between the amount of improvement on the LACE Speech-in-Noise task and the pretraining strength of pitch encoding during the formant transition ($r = -0.394$, $P = 0.038$) and steady-state regions ($r = -0.435$, $P = 0.021$). A similar relationship was found between the Competing Speaker and pitch encoding during the steady-state region ($r = -0.467$, $P = 0.012$), suggesting that the strength of an individual person's subcortical encoding of pitch cues might predict the capacity for improvement with this particular training program.

Improvement on independent measures of speech-in-noise perception also correlated significantly with pretest pitch-related brainstem measures. Pearson r correlations were statistically significant between the improvement in QuickSIN loss scores and the composite pitch measure obtained from both the transition and steady-state regions prior to training ($r = -0.497$, $P = 0.007$; $r = -0.580$, $P = 0.001$, respectively). Moreover, we found a significant relationship with the improvement in performance on HINT-Front and this neural measure during the steady-state region ($r = -0.552$, $P = 0.002$). Thus, participants with the largest strength of pitch encoding (F_0 and H_2 magnitude) at pretest showed the greatest amount of generalization of learning to independent measures of speech-in-noise perception (Fig. 6), consistent with their greater learning on the training measures.

In summary, we found training-related enhancements in speech-in-noise perception (LACE tasks, QuickSIN, and HINT) and increased neural representation of pitch cues in background noise. The neural enhancement in noise was most pronounced during the time-varying portion of the syllable known to be most vulnerable to perceptual disruption (i.e., the transition region).

Discussion

This study is the first to show that the biological mechanisms subserving speech perception in noise are malleable with short-term training, and that the benefits of training are maintained over time. While training-induced brainstem plasticity has been demonstrated previously (Russo et al. 2005; Song et al. 2008; Carcagno and Plack 2010), unlike its predecessors, we utilized a more naturalistic training regimen that targets hearing speech in noise. By showing that a general

training approach that mimics the challenges of real-world listening can affect sound processing in the brainstem, we provide an important conceptual advance to our understanding of the kind of training experiences that can influence sensory processing in adulthood.

Cognitive Influences on Subcortical Plasticity

To examine neural changes associated with improved speech perception in noise, we employed a commercially available program (LACE) that incorporates many features that maximize auditory learning (i.e., feedback, positive reinforcement, a diverse stimulus set, activation of cognitive processes). Due to the nature of the training program and the fact that improvements were seen on all LACE exercises, we cannot tease out whether subcortical plasticity was driven by a specific listening exercise or by the cumulative effects of all exercises. However, we contend that changes in perception and neurophysiology likely resulted from the way in which LACE integrates cognitive factors into its auditory training exercises. This kind of approach is especially important for improving our ability to listen in noise because sensory and cognitive processes must operate in tandem to extract and decode the target signal. Thus, by invoking high cognitive demands (e.g., semantically and syntactically complex sentences, auditory working memory, and attention) (Sweetow et al. 2006), LACE may strengthen cortical processes which in turn improve sensory acuity in the brainstem when listening to speech in background noise. As a consequence of this feedback, the signal being relayed to the cortex is more robust. This explanation is consistent with a theoretical framework of subcortical plasticity in which the brainstem operates as part of an integrated network of subcortical and cortical structures linked by afferent and efferent processes (Suga et al. 2000; Krishnan and Gandour 2009; Tzounopoulos and Kraus 2009; Bajo et al. 2010; Kraus and Chandrasekaran 2010). In this theoretical framework, brainstem activity can both "influence" and "be influenced by" cortical processes. Here, we provide evidence that this network is malleable with training, by showing that improved brainstem encoding can arise from training on cognitive-based tasks.

It is important to note that this short-term training program did not have broad indiscriminate effects on auditory processing in noise. Instead, it affected how specific aspects of the auditory signal, namely pitch cues (F_0 and H_2), were processed.
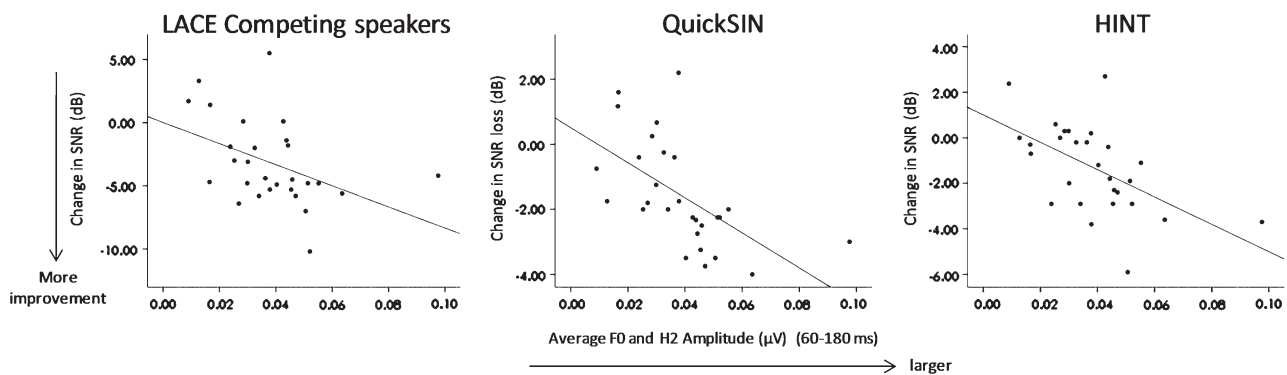
Figure 6. Correlation between behavioral and physiologic measures. Training-related changes in speech-in-noise performance (posttraining minus pretraining) were plotted as a function of the strength of pitch encoding (average magnitude of F_0 and H_2) obtained from the steady-state region in the 6-talker babble condition at pretest. Pearson r correlations were significant for LACE-related tasks involving background noise (Competing speaker; $r = -0.467$, $P = 0.012$) and the degree of improvement on QuickSIN ($r = -0.580$, $P = 0.001$) and HINT Noise-Front ($r = -0.552$, $P = 0.002$).

How can we explain this result? From an acoustic standpoint, F_0 and H_2 are important cues when listening in noise because they aid speaker identification and auditory object formation (Oxenham 2008; Shinn-Cunningham and Best 2008). As suggested by previous work (Song et al. 2008; Anderson et al. 2010; Anderson et al. 2011), the ability to understand speech in noise is influenced by how the subcortical auditory system represents these specific pitch cues. It is also known that the time-variant (formant transition) portion of the syllable is especially perceptually vulnerable in noise (Miller and Nicely 1955; Assmann and Summerfield 2004). Notably, the training-associated changes reported here were reflected specifically in the encoding of F_0 and H_2 and were most salient in response to the transition portion of the syllable. Here, we add a new piece to the story by showing that pitch encoding can be improved when training under naturalistic conditions in which pitch cues are perceptually important.

In summary, improvements in sensory processing in noise likely resulted from the interaction of the brainstem's "affinity" for encoding pitch cues (Krishnan 2002; Krishnan and Gandour 2009; Krishnan et al. 2009), the inherent malleability of pitch processing (Wong, Skoe, et al. 2007; Song et al. 2008; Carcagno and Plack 2010), and the importance of pitch cues when listening in noise. Moreover, LACE's cognitive approach to training in noise may have boosted the effects of training by engaging top–down mechanisms.

Predicting Learning Success: Behavioral and Neural Metrics

Consistent with de Boer and Thornton's (2008) work, learning was dependent on the initial performance on a measure of speech-in-noise perception (i.e., QuickSIN), such that poorer performers showed greater improvements than good performers. We also show that the extent of learning was predicted by the strength of the neural representation of pitch cues presented in noise before training. This finding suggests that individuals, who are better at representing pitch cues before training, are in a stronger position to profit from these cues in the training exercises, which might enable greater improvements in speech-in-noise perception. This interpretation is consistent with cortical studies in which the most successful learners were those with the largest pretraining activation (Wong, Perrachione, et al. 2007; Tremblay et al. 2009). Future investigations will help to further delineate the complex interactions among the behavioral and neurophysiological factors that predict learning outcomes.

Clinical and Social Implications

By showing that subcortical function is malleable in adulthood, our study corroborates and extends findings by de Boer and Thornton (2008), Song et al. (2008), and Carcagno and Plack (2010). The inclusion of a young, adult, normal hearing population allowed us to elucidate the nervous system's capacity to change without the confounds of age-related declines in peripheral and cognitive functions. While nonnative English speakers are not a clinical group, they are an important population to study for both scientific and sociopolitical reasons. For example, they experience greater difficulty perceiving their second language in noise despite having normal hearing and intelligence (Mayo et al. 1997; Meador et al. 2000; Rogers et al. 2006). This has significant sociopolitical

implications given that the breadth of linguistic diversity continues to increase rapidly in the United States of America. Over the past decade, numerous states have experienced double-digit population growth rates largely due to a greater influx of nonnative English-speaking immigrants. Approximately 11% (31 million) of individuals counted in the 2000 US Census were not born in this country (United States Census 2000), and recently available data from the 2004–2005 school year reported that approximately 5.1 million nonnative English learners were enrolled in public schools, a 45% increase over the 1998–1999 academic year (NCELA 2006). The pronounced increase of nonnative English speakers in the United States of America should motivate researchers to investigate avenues to enhance the effectiveness of communication in everyday listening situations.

Our study supports the use of cognitive-based auditory training programs to improve listening in everyday listening environments for normal hearing young adults. Thus, our findings provide an important baseline for understanding the biological correlates of training-related plasticity in adulthood. Taken together with the fact that LACE was designed for clinical populations with age-related changes in auditory function and those with peripheral hearing loss, findings from this study and others could be used to ultimately enhance human communication for those experiencing greater auditory perceptual difficulties, such as older adults (Kim et al. 2006; Wong et al. 2010; Parbery-Clark et al. 2011), individuals with hearing loss (Dubno et al. 1984; Helfer and Wilber 1990), and children with learning problems (Bradlow et al. 2003; Ziegler et al. 2005, 2009).

Specifically, we found that LACE training generalizes to standardized clinically utilized measures of speech-in-noise perception—a critical factor if training is to have an impact on real-world listening. This independent assessment of a commercially available training program has the potential to provide valuable information to clinicians, teachers, and the general public. This study is a step toward verifying the efficacy of auditory training programs, which, together with other studies investigating training-related improvements in speech perception, will help empower clinicians to optimize remediation strategies for populations with communication difficulties. Moreover, by contributing to our understanding of the neural mechanisms subserving to auditory learning, our results may prove helpful in designing intervention protocols for improving speech-in-noise perception in everyday listening situations.

Future investigations should extend this line of work to a wider range of populations. Since hearing in noise is fundamental to human communication, the outcomes of this future work could have broad social implications. This is because the failure to effectively communicate in noise can impede vocational, emotional, and academic success.

Conclusion

This study offers the first empirical evidence that naturalistic training, combining sensory and cognitive dimensions, can enhance how the nervous system encodes acoustic cues that are important for listening in noise. This finding, which provides a conceptual advance to our understanding of experience-dependent plasticity, aligns with a theoretical framework, in which subcortical and cortical structures work in a cohesive manner to enable complex auditory learning

(Suga et al. 2000; Krishnan and Gandour 2009; Bajo et al. 2010; Kraus and Chandrasekaran 2010). By showing that a training program can improve how the adult nervous system processes speech in noise, these results promote the possibility of improving a fundamental, yet challenging, aspect of human communication in a wide range of populations.

References

Aiken SJ, Picton TW. 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear Res. 245:35–47.

Akhoun I, Gallégo S, Moulin A, Menard M, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H. 2008. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. Clin Neurophysiol. 119:922–933.

Amitay S, Hawkey DJC, Moore DR. 2005. Auditory frequency discrimination learning is affected by stimulus variability. Percept Psychophys. 67:691–698.

Ananthanarayan AK, Durrant JD. 1992. The frequency-following response and the onset response: evaluation of frequency specificity using a forward-masking paradigm. Ear Hear. 13:228–232.

Anderson S, Parbery-Clark A, Yi HG, Kraus N. 2011. A neural basis of speech-in-noise perception in older adults. Ear Hear. 32. doi: 10.1097/AUD.0b013e31822229d3.

Anderson S, Skoe E, Chandrasekaran B, Zecker S, Kraus N. 2010. Brainstem correlates of speech-in-noise perception in children. Hear Res. doi: 10.1016/j.heares.2010.08.001.

Assmann P, Summerfield Q. 2004. The perception of speech under adverse conditions. New York: Springer.

Assmann PF, Summerfield Q. 1990. Modeling the perception of concurrent vowels—vowels with different fundamental frequencies. J Acoust Soc Am. 88:680–697.

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

Bakin JS, Weinberger NM. 1990. Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. Brain Res. 536:271–286.

Bidelman GM, Krishnan A. 2010. Effects of reverberation on brainstem representation of speech in musicians and nonmusicians. Brain Res. 1355:112–125.

Bird J, Darwin C. 1998. Effects of a difference in fundamental frequency in separating two sentences. Psychophysical and physiological advances in hearing. London: Whurr.

Bradlow AR, Kraus N, Hayes E. 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear Res. 46:80–97.

Brokx JPL, Nooteboom SG. 1982. Intonation and the perceptual separation of simultaneous voices. J Phon. 10:23–36.

Brown L, Sherbenou RJ, Johnsen SK. 1997. Test of nonverbal intelligence: a language-free measure of cognitive ability. Austin (TX): PRO-ED Inc.

Burk MH, Humes LE. 2007. Effects of training on speech recognition performance in noise using lexically hard words. J Speech Lang Hear Res. 50:25–40.

Burk MH, Humes LE, Amos NE, Strauser LE. 2006. Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. Ear Hear. 27:263–278.

Cainer KE, James C, Rajan R. 2008. Learning speech-in-noise discrimination in adult humans. Hear Res. 238:155–164.

Carcagno S, Plack C. 2010. Subcortical plasticity following perceptual learning in a pitch discrimination task. J Assoc Res Otolaryngol. doi: 10.1007/s10162-010-0236-1.

Chandrasekaran B, Kraus N. 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. Psychophysiology. 47:236–246.

de Boer J, Thornton ARD. 2008. Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. J Neurosci. 28:4929–4937.

Diamond DM, Weinberger NM. 1984. Physiological plasticity of single neurons in auditory cortex of the cat during acquisition of the pupillary conditioned response: II. Secondary field (AII). Behav Neurosci. 98:189–210.

Diamond DM, Weinberger NM. 1986. Classical conditioning rapidly induces specific changes in frequency receptive fields of single neurons in secondary and ventral ectosylvian auditory cortical fields. Brain Res. 372:357–360.

Diamond DM, Weinberger NM. 1989. Role of context in the expression of learning-induced plasticity of single neurons in auditory cortex. Behav Neurosci. 103:471–494.

Dubno JR, Dirks DD, Morgan DE. 1984. Effects of age and mild hearing-loss on speech recognition in noise. J Acoust Soc Am. 76:87–96.

Edeline JM, Pham P, Weinberger NM. 1993. Rapid development of learning-induced receptive field plasticity in the auditory cortex. Behav Neurosci. 107:539–551.

Gaab N, Gaser C, Schlaug G. 2006. Improvement-related functional plasticity following pitch memory training. Neuroimage. 31:255–263.

Galbraith GC, Amaya EM, de Rivera JM, Donan NM, Duong MT, Hsu JN, Tran K, Tsang LP. 2004. Brain stem evoked response to forward and reversed speech in humans. Neuroreport. 15:2057–2060.

Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM. 1995. Intelligible speech encoded in the human brain stem frequency following response. Neuroreport. 6:2363–2367.

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

Gorga M, Abbas P, Worthington D. 1985. Stimulus calibration in ABR measurements. In: Jacobsen J, editor. The auditory brainstem response. San Diego (CA): College Hill Press. p. 49–62.

Hall JW. 1992. Handbook of auditory evoked responses. Needham Heights (MA): Allyn and Bacon.

Hedrick MS, Younger MS. 2007. Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise. J Speech Lang Hear Res. 50:254–269.

Helfer KS, Wilber LA. 1990. Hearing-loss, aging, and speech-perception in reverberation and noise. J Speech Hear Res. 33:149–155.

Hood LJ. 1998. Clinical applications of the auditory brainstem response. San Diego (CA): Singular Publishing Group, Inc.

Jacobson J. 1985. The auditory brainstem response. San Diego (CA): College-Hill Press.

Johnston KN, John AB, Kreisman NV, Hall JW, Crandell CC. 2009. Multiple benefits of personal FM system use by children with auditory processing disorder (APD). Int J Audiol. 48:371–383.

Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. 2004. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J Acoust Soc of Am. 116:2395–2405.

Kim SH, Frisina RD, Mapes FM, Hickman ED, Frisina DR. 2006. Effect of age on binaural speech intelligibility in normal hearing adults. Speech Commun. 48:591–597.

King C, Warrier CM, Hayes E, Kraus N. 2002. Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neurosci Lett. 319:111–115.

Klatt D. 1980. Software for cascade/parallel formant synthesizer. J Acoust Soc Am. 67:971–975.

Kraus N, Chandrasekaran B. 2010. Musical training for the development of auditory skills. Nat Rev Neurosci. 11:599–605.

Kraus N, Disterhoft JF. 1982. Response plasticity of single neurons in rabbit auditory association cortex during tone-signalled learning. Brain Res. 246:205–215.

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

Kraus N, Nicol T. 2005. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci. 28:176–181.

Krishnan A. 2002. Human frequency-following responses: representation of steady-state synthetic vowels. Hear Res. 166:192–201.

Krishnan A, Gandour JT. 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang. 110:135–148.

Krishnan A, Gandour JT, Bidelman GM. 2010. Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. Brain Res. 1313:124–133.

Krishnan A, Gandour JT, Bidelman GM, Swaminathan J. 2009. Experience-dependent neural representation of dynamic pitch in the brainstem. Neuroreport. 20:408–413.

Krishnan A, Xu Y, Gandour JT, Cariani PA. 2004. Human frequency-following response: representation of pitch contours in Chinese tones. Hear Res. 189:1–12.

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

Kujala T, Karma K, Ceponiene R, Belitz S, Turkkila P, Tervaniemi M, Naatanen R. 2001. Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. Proc Natl Acad Sci U S A. 98:10509–10514.

Kumar A, Hedge M, Mayaleela. 2010. Perceptual learning of non-native speech contrast and functioning of the olivocochlear bundle. Int J Audiol. 49:488–496.

Li X, Jeng FC. 2011. Noise tolerance in human frequency-following responses to voice pitch. J Acoust Soc Am. 129:EL21–EL26.

Mayo LH, Florentine M, Buus S. 1997. Age of second-language acquisition and perception of speech in noise. J Speech Lang Hear Res. 40:686–693.

Meador D, Flege JE, Mackay IR. 2000. Factors affecting the recognition of words in a second language. Lang Cogn. 3:55–67.

Meddis R, O'Mard L. 1997. A unitary model of pitch perception. J Acoust Soc Am. 102:1811–1820.

Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. 1996. Temporal processing deficits of language-learning impaired children ameliorated by training. Science. 271:77–81.

Miller GA, Nicely PE. 1955. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am. 27:338–352.

Møller AR, Jannetta P. 1985. Neural generators of the auditory brainstem response. In: Jacobsen J, editor. The auditory brainstem response. San Diego (CA): College Hill Press. p. 13–32.

Moore DR, Amitay S. 2007. Auditory training: rules and applications. Semin Hear. 28:99–109.

Moore DR, Rosenberg JF, Coleman JS. 2005. Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. Brain Lang. 94:72–85.

Mossbridge JA, Fitzgerald MB, O'Connor ES, Wright BA. 2006. Perceptual-learning evidence for separate processing of asynchrony and order tasks. J Neurosci. 26:12708–12716.

Moushegian G, Rupert AL, Stillman RD. 1973. Scalp-recorded early responses in man to frequencies in speech range. Electroencephalogr Clin Neurophysiol. 35:665–667.

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

Naatanen R, Schroger E, Karakas S, Tervaniemi M, Paavilainen P. 1993. Development of a memory trace for a complex sound in the human brain. Neuroreport. 4:503–506.

Nabelek AK, Donahue AM. 1984. Perception of consonants in reverberation by native and non-native listeners. J Acoust Soc Am. 75:632–634.

NCELA 2006. Elementary and secondary enrollment of ELL students in U.S. Washington (DC): National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs Available from: URL http://www.ncela.gwu.edu/expert/faq/08leps.html.

Neff DL, Green DM. 1987. Masking produced by spectral uncertainty with multicomponent maskers. Percept Psychophys. 41:409–415.

Nilsson M, Soli SD, Sullivan JA. 1994. Development of the Hearing in Noise Test for the Measurement of Speech Reception Thresholds in Quiet and in Noise. J Acoust Soc of Am. 95:1085–1099.

Olds J, Disterhoft JF, Segal M, Kornblith CL, Hirsh R. 1972. Learning centers of rat brain mapped by measuring latencies of conditioned unit responses. J Neurophysiol. 35:202–219.

Oxenham AJ. 2008. Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. Trends Amplif. 12:316–331.

Parbery-Clark A, Strait DL, Anderson S, Hittner E, Kraus N. 2011. Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. PLoS ONE 6:e18082. doi: 10.1371/journal.pone.0018082.

Parbery-Clark A, Skoe E, Kraus N. 2009b. Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci. 29:14100–14107.

Parbery-Clark A, Skoe E, Lam C, Kraus N. 2009a. Musician enhancement for speech in noise. Ear Hear. 30:653–661.

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

Rogers CL, Lister JJ, Febo DM, Besing JM, Abrams HB. 2006. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. Appl Psycholinguist. 27:465–485.

Russo N, Nicol T, Musacchia G, Kraus N. 2004. Brainstem responses to speech syllables. Clin Neurophysiol. 115:2021–2030.

Russo N, Nicol T, Trommer B, Zecker S, Kraus N. 2009. Brainstem transcription of speech is disrupted in children with autism spectrum disorders. Dev Sci. 12:557–567.

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

Schaffler T, Sonntag J, Hartnegg K, Fischer B. 2004. The effect of practice on low-level auditory discrimination, phonological skills, and spelling in dyslexia. Dyslexia. 10:119–130.

Shinn-Cunningham BG, Best V. 2008. Selective attention in normal and impaired hearing. Trends Amplif. 12:283–299.

Skoe E, Kraus N. 2010. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 31:302–324.

Smiljanic R, Bradlow AR. 2005. Production and perception of clear speech in Croatian and English. J Acoust Soc Am. 118:1677–1688.

Smith GE, Housen P, Yaffe K, Ronald R, Kennison RF, Mahncke HW, Zelinski EM. 2009. A cognitive training program based on principles of brain plasticity: results from the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) study. J Am Geriatr Soc. 57:594–603.

Song JH, Nicol T, Kraus N. 2010a. Test-retest reliability of the speech-evoked auditory brainstem response in young adults. Clin Neurophysiol. doi: 10.1016/j.clinph.2010.07.009.

Song JH, Skoe E, Banai K, Kraus N. 2010b. Perception of speech in noise: neural correlates. J Cogn Neurosci. doi: 10.1162/jocn.2010.21556.

Song JH, Skoe E, Wong PCM, Kraus N. 2008. Plasticity in the adult human auditory brainstem following short-term linguistic training. J Cogn Neurosci. 20:1892–1902.

Strait DL, Kraus N, Skoe E, Ashley R. 2009. Musical experience and neural efficiency—effects of training on subcortical processing of vocal expressions of emotion. Eur J Neurosci. 29:661–668.

Suga N, Gao EQ, Zhang YF, Ma XF, Olsen JF. 2000. The corticofugal system for hearing: recent progress. Proc Natl Acad Sci U S A. 97:11807–11814.

Swaminathan J, Krishnan A, Gandour JT. 2008. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. Neuroreport. 19:1163–1167.

Sweetow RW, Sabes JH. 2006. The need for and development of an adaptive Listening and Communication Enhancement (LACE) Program. J Am Acad Audiol. 17:538–558.

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

Tallal P, Piercy M. 1974. Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. Neuropsychologia. 12:83–93.

Tice R, Carrell T. 1998. Level 16. (Version 2.0.3) [Computer Software]. Lincoln (NE): University of Nebraska.

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

Tremblay KL, Shahin AJ, Picton T, Ross B. 2009. Auditory training alters the physiological detection of stimulus-specific cues in humans. Clin Neurophysiol. 120:128–135.

Tzounopoulos T, Kraus N. 2009. Learning to encode timing: mechanisms of plasticity in the auditory brainstem. Neuron. 62:463–469.

United States Bureau of the Census. 2000. Census of Demographic Profiles. http://www.census.gov/Pres-Release/www.2002/demo-profiles.html. Washington (DC): The Bureau.

Van Engen KJ, Bradlow AR. 2007. Sentence recognition in native- and foreign-language multi-talker background noise. J Acoust Soc Am. 121:519–526.

Weinberger NM. 1993. Learning-induced changes of auditory receptive fields. Curr Opin Neurobiol. 3:570–577.

Wilson JR, Krishnan A. 2005. Human frequency-following responses to binaural masking level difference stimuli. J Am Acad Audiol. 16:184–195.

Wilson RH, McArdle RA, Smith SL. 2007. An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. J Speech Lang Hear Res. 50:844–856.

Wong PCM, Ettlinger M, Sheppard JP, Gunasekera GM, Dhar S. 2010. Neuroanatomical characteristics and speech perception in noise in older adults. Ear Hear. 31:471–479.

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

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

Wright BA, Buonomano DV, Mahncke HW, Merzenich MM. 1997. Learning and generalization of auditory temporal-interval discrimination in humans. J Neurosci. 17:3956–3963.

Xu YS, Krishnan A, Gandour JT. 2006. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport. 17:1601–1605.

Yamada O, Kodera K, Hink FR, Suzuki JI. 1979. Cochlear distribution of frequency-following response initiation. A high-pass masking noise study. Audiology. 18:381–387.

Yund EW, Woods DL. 2010. Content and procedural learning in repeated sentence tests of speech perception. Ear Hear. 31:769–778.

Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C. 2005. Deficits in speech perception predict language learning impairment. Proc Natl Acad Sci U S A. 102:14110–14115.

Ziegler JC, Pech-Georgel C, George F, Lorenzi C. 2009. Speech-perception-in-noise deficits in dyslexia. Dev Sci. 12:732–745.