Perception of Speech in Noise: Neural Correlates

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

Abstract

■ The presence of irrelevant auditory information (other talkers, environmental noises) presents a major challenge to listening to speech. The fundamental frequency (F_0) of the target speaker is thought to provide an important cue for the extraction of the speaker's voice from background noise, but little is known about the relationship between speech-in-noise (SIN) perceptual ability and neural encoding of the F_0. Motivated by recent findings that music and language experience enhance brainstem representation of sound, we examined the hypothesis that brainstem encoding of the F_0 is diminished to a greater degree by background noise in people with poorer perceptual abilities in noise. To this end, we measured speech-evoked auditory brainstem responses to /da/ in quiet and two multi-talker babble conditions (two-talker and six-talker) in native English-speaking young adults who ranged in their ability to perceive and recall SIN. Listeners who were poorer performers on a standardized SIN measure demonstrated greater susceptibility to the degradative effects of noise on the neural encoding of the F_0. Particularly diminished was their phase-locked activity to the fundamental frequency in the portion of the syllable known to be most vulnerable to perceptual disruption (i.e., the formant transition period). Our findings suggest that the subcortical representation of the F_0 in noise contributes to the perception of speech in noisy conditions. ■

INTRODUCTION

Extracting a speaker's voice from a background of competing voices is essential to the communication process. This process is often challenging, even for young adults with normal hearing and cognitive abilities (Assmann & Summerfield, 2004; Neff & Green, 1987). Successful extraction of the target message is dependent on the listener's ability to benefit from speech cues, including the fundamental frequency (F_0) of the target speaker (Stickney, Assmann, Chang, & Zeng, 2007; Summers & Leek, 1998; Brokx & Nooteboom, 1982). The F_0 of speech sounds is an important acoustic cue for speech perception in noise because it allows for grouping of speech components across frequency and over time (Bird & Darwin, 1998; Brokx & Nooteboom, 1982), which aids in speaker identification (Baumann & Belin, 2010). Although the perceptual ability to track the F_0 plays a crucial role in speech-in-noise (SIN) perception, the relationship between SIN perceptual ability and neural encoding of the F_0 has not been established. Thus, we aimed to determine whether there is a relationship between SIN perception and neural representation of the F_0 in the brainstem. To that end, we examined the phase-locked activity to the fundamental periodicity of the auditory brainstem response (ABR) to the speech syllable /da/ presented in quiet and background noise in native English-speaking young adults who ranged in their ability to perceive SIN.

The frequency following response (FFR), which reflects phase-locked activity elicited by periodic acoustic stimuli, is produced by populations of neurons along the auditory brainstem pathway (Chandrasekaran & Kraus, 2010). This information is preserved in the interpeak intervals of the FFR, which are synchronized to the period of the F_0 and its harmonics. The FFR has been elicited by a wide range of periodic stimuli, including pure tones and masked tones (Marler & Champlin, 2005; McAnally & Stein, 1997), English and Mandarin speech syllables (Aiken & Picton, 2008; Akhoun et al., 2008; Swaminathan, Krishnan, & Gandour, 2008; Wong, Skoe, Russo, Dees, & Kraus, 2007; Xu, Krishnan, & Gandour, 2006; Krishnan, Xu, Gandour, & Cariani, 2004, 2005; Galbraith et al., 2004; Russo, Nicol, Musacchia, & Kraus, 2004; King, Warrier, Hayes, & Kraus, 2002; Krishnan, 2002), words (Wang, Nicol, Skoe, Sams, & Kraus, 2009; Galbraith et al., 2004), musical notes (Bidelman, Gandour, & Krishnan, 2011; Lee, Skoe, Kraus, & Ashley, 2009; Musacchia, Sams, Skoe, & Kraus, 2007), and emotionally valent vocal sounds (Strait, Kraus, Skoe, & Ashley, 2009), mostly presented in a quiet background. The FFR reflects specific spectral and temporal properties of the signal, including the F_0, with such precision that the response can be recognized as intelligible speech when "played back" as an auditory stimulus (Galbraith, Arbagey, Branski, Comerci, & Rector, 1995). This allows comparisons between the response frequency composition and the corresponding features of the stimulus (Skoe & Krauss, 2010; Kraus & Nicol, 2005; Russo et al., 2004; Galbraith et al., 2000). Thus, the FFR offers an objective measure of the degree to which the auditory pathway accurately encodes complex acoustic

[1]Northwestern University, [2]University of Haifa

features, including those known to play a role in characterizing speech under adverse listening conditions. The FFR is also sensitive to the masking effects of competing sounds (Anderson, Chandrasekaran, Skoe, & Kraus, 2010; Parbery-Clark, Skoe, & Kraus, 2009; Russo, Nicol, Trommer, Zecker, & Kraus, 2009; Wilson & Krishnan, 2005; Russo et al., 2004; Ananthanarayan & Durrant, 1992; Yamada, Kodera, Hink, & Suzuki, 1979), resulting in delayed and diminished responses (i.e., neural asynchrony).

A relationship exists between perceptual abilities, listener experience, and brainstem encoding of complex sounds. For example, lifelong experience with linguistic F_0 contours, such as those occurring in Mandarin Chinese, enhances the subcortical F_0 representation (Krishnan, Swaminathan, & Gandour, 2009; Swaminathan et al., 2008; Krishnan et al., 2005). Similarly, musicians have more robust F_0 encoding for speech sounds (Kraus & Chandrasekaran, 2010; Musacchia et al., 2007), nonnative linguistic F_0 contours (Wong et al., 2007), and emotionally salient vocal sounds (Strait et al., 2009) compared with nonmusicians. Enhancement of brainstem activity accompanies short-term auditory training (deBoer & Thorton, 2008; Russo, Nicol, Zecker, Hayes, & Kraus, 2005), including training-related increased accuracy of F_0 encoding (Song, Skoe, Wong, & Kraus, 2008). Moreover, the neural representation of F_0 contours is diminished in a subset of children with autism spectrum disorders who have difficulties understanding speech prosody (Russo et al., 2008). Relevant to the present investigation, speech perception in noise is related to the subcortical encoding of stop-consonant stimuli (Anderson et al., 2010; Hornickel, Skoe, Nicol, Zecker, & Kraus, 2009; Parbery-Clark et al., 2009; Tzounopoulos & Kraus, 2009) as well as the effectiveness of the nervous system to extract regularities in speech sounds relevant for vocal pitch (Chandrasekaran, Hornickel, Skoe, Nicol, & Kraus, 2009). These findings suggest that the representation of the F_0 and other components of speech in the brainstem are related to specific perceptual abilities.

In this study, we sought to determine whether SIN perception is associated with neural representation of the F_0 in the brainstem as well as advance our understanding of the subcortical encoding of speech sounds in multi-talker noise. Although previous studies have investigated the impact of background noise on brainstem responses to speech in children using nonspeech background noise (i.e., white Gaussian noise; Russo et al., 2009; Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2001), multitalker babble was chosen as the background noise because it closely resembles naturally occurring listening conditions where listeners are required to extract the voice of the target speaker from a background of competing voices. Using two- and six-talker babble, we investigated the degree to which different levels of energetic masking affect subcortical encoding of speech. In contrast to the two-talker babble, the six-talker babble imparts nearly full energetic masking (i.e., fewer spectral and temporal gaps than the two-talker babble) and thus constitutes a more challenging listening environment.

We examined the effect of noise on the strength of F_0 encoding within the regions of the response thought to correspond to the formant transition and steady-state vowel separately. The formant transition region of the syllable poses particular perceptual challenges even for normal hearing listeners (Assmann & Summerfield, 2004; Miller & Nicely, 1955). Importantly, although the F_0 was constant in terms of frequency and amplitude throughout the syllable (see Figure 1B), within the formant transition region the phase of the F_0 is more variable because of the influences of the rapidly changing formant frequencies (see Figure 1C). In contrast, the F_0 of the steady-state vowel is reinforced by the unwavering formants that fall at integer multiples of the F_0. Given that the formant transition is known to have greater perceptual susceptibility in noise and because a weaker (i.e., more variable) F_0 cue imposes larger neural synchronization demands on brainstem neurons (Krishnan, Gandour, & Bidelman, 2010), we hypothesized that noise would have a greater impact on the phase-locked activity to the fundamental periodicity of this region compared with the steady-state vowel. Thus, if a neural system is sensitive to the effects of desynchronization, this susceptibility would become apparent in noise and manifest as less robust representation of the F_0 relative to the response in quiet.

METHODS

Participants

Seventeen monolingual native English-speaking adults (13 women, age = 20–31 years, mean age = 24 years, SD = 3 years) with no history of neurological disorders participated in this study. To control for musicianship, a factor known to modulate F_0 encoding at the level of the brainstem, all subjects had fewer than 6 years of musical training that ceased 10 or more years before study enrollment. All participants had normal IQ (mean = 107.5, SD = 8), as measured by the Test of Nonverbal Intelligence 3 (Brown, Sherbenou, & Johnsen, 1997), normal hearing (≤20 dB HL pure-tone thresholds from 125 to 8000 Hz), and normal click-evoked auditory response wave V latencies to 100-μs clicks presented at 31.1 times per second 80.3 dB SPL. Participants gave their informed consent in accordance with the Northwestern University institutional review board regulations.

Behavioral Procedures and Analyses

Quick Speech-in-Noise Test (QuickSIN; Etymotic Research, Elk Grove Village, IL; Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004) is a nonadaptive test of speech perception in four-talker babble (three women and one man) that is commonly used during audiologic testing to assess
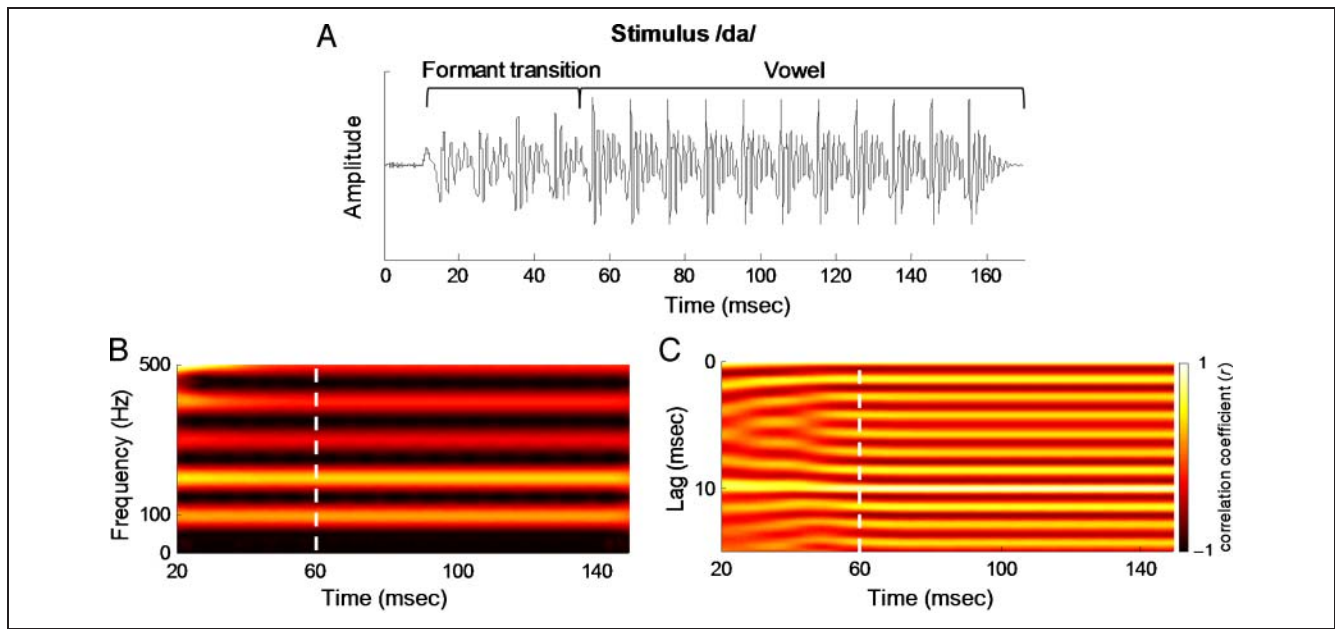
Figure 1. Stimulus characteristics. (A) The acoustic waveform of the target stimulus /da/. The formant transition and the vowel regions are bracketed. The periodic amplitude modulations of the stimulus, reflecting the rate of the fundamental frequency, are represented by the major peaks in the stimulus waveform (10 msec apart). (B) The spectrogram illustrating the fundamental frequency and lower harmonics (stronger amplitudes represented with brighter colors) and (C) the autocorrelogram (a visual measure of response periodicity) of the stimulus /da/. The boundary of the consonant-vowel formant transition and the steady-state vowel portion of the syllable is marked by a dashed white line. Although the frequency and spectral amplitude of the F_0 are constant as shown by the spectrogram, the interaction of the formants with the F_0 in our stimulus resulted in weaker fundamental periodicity in the formant transition period (more diffuse colors). In contrast, the vowel is composed of unchanging formants, resulting in sustained and stronger F_0 periodicity as shown by the autocorrelogram. These plots were generated via running window analysis over 40-msec bins starting at time 0, and the x axis refers to the midpoint of each bin (Song et al., 2008).

speech perception in noise in adults. QuickSIN was presented binaurally to participants through insert earphones (ER-2; Etymotic Research). The test is composed of 12 lists of sentences. Each list consists of six sentences spoken by a single adult female speaker, with five target words per sentence. Each participant was given one practice list to acquaint him or her with the task. Then 4 of the 12 lists of sentences were randomly selected and administered. Participants were instructed to repeat back each sentence. The sentences were syntactically correct, but the target words were difficult to predict because of limited semantic cues (Wilson, McArdle, & Smith, 2007); e.g., target words are italicized: *Dots* of *light betrayed* the *black cat*. With each subsequent sentence, the signal-to-noise ratio (SNR) became increasingly more difficult as a result of the intensity of the background babble being held constant at 70 dB HL, and the intensity of the target sentence decreasing. The first sentence was presented at a SNR of 25 dB and then the SNR decreased in 5-dB steps with each subsequent sentence so that the last sentence was presented at 0-dB SNR. In the +5 SNR condition, eight subjects performed at 100%, and no subject scored lower than 75%. To avoid this ceiling effect, we based our measure of SIN performance in this study on the percentage of target words out of 20 correctly recalled under the most challenging SNR (0 dB). Participants were placed into two groups on the basis of the median SIN performance score, which was 25% (5 of 20 correct words). Those with a SIN perfor-

mance score equal to or higher than the median score were placed in the "top" SIN group ($n = 9$). Participants with a SIN performance scores lower than the median score were grouped into the "bottom" SIN group ($n = 8$).

Sentence recognition threshold in quiet was measured using the Hearing in Noise Test (Bio-logic Systems Corp., Mundelein, IL; Nilsson, Soli, & Sullivan, 1994). Top and bottom SIN groups showed similar sentence recognition thresholds in quiet (mean = 20.75 dB, SD = 2.35 dB and mean = 20.83, SD = 3.34 dB, respectively), $t = -.06, p = .953$. This result suggested that both groups had comparable ability to perceive speech in a quiet listening environment, making it unlikely that any difference observed in noise was the result of one group having overall better speech perception capabilities rather than better perception of SIN.

Neurophysiologic Stimuli and Design

Brainstem responses were elicited in response to the syllable /da/ in quiet and two noise conditions. /Da/ is a five-formant syllable synthesized at a 20-kHz sampling rate using a Klatt synthesizer. The duration was 170 msec, with a voicing (100 Hz F_0) onset at 10 msec (see Figure 1A and B). Formant transition duration was 50 msec and comprised a linearly rising first formant (400–720 Hz), linearly falling second and third formants (1700–1240 and 2580–2500 Hz, respectively), and flat fourth (3300 Hz), fifth

(3750 Hz), and sixth formants (4900 Hz). After the transition period, these formant frequencies remained constant at 720, 1240, 2500, 3300, 3750, and 4900 Hz for the remainder of the syllable. The stop burst consisted of 10 msec of initial frication centered at frequencies around F_4 and F_5. The syllable /da/ was presented in alternative polarities via a magnetically shielded insert earphone placed in the right ear (ER-3; Etymotic Research) at 80.3 dB SPL at a rate of 4.35 Hz.

The noise conditions consisted of multitalker babble spoken in English. Two-talker babble (one woman and one man, 20-sec track) and six-talker babble (three women and three men, 4-sec track) were selected because they provide different levels of energetic masking. To create the babble, we instructed the speakers to speak in a natural, conversational style. Recordings were made in a sound-attenuated booth in the phonetics laboratory of the Department of Linguistics at Northwestern University for unrelated research (Smiljanic & Bradlow, 2005) and were digitized at a sampling rate of 16 kHz with 24-bit accuracy (for further details, see Van Engen & Bradlow, 2007; Smiljanic & Bradlow, 2005). The tracks were root mean square amplitude normalized using Level 16 software (Tice & Carrell, 1998). To create the multitalker babble, we staggered the start time of each speaker by inserting silence (100–500 msec) to the beginning of five of the six speakers' babble tracks (layers), mixing all six babble layers to become one track and then trimming off the first 500 msec. This procedure led to sections of each layer being removed: the first 500 msec of the first layer, 400 msec of the second layer, the first 300 msec of the third layer, first 200 msec of the fourth layer, and the first 100 msec of the fifth layer. The two layers of the two-babble track were also staggered by adding 500 msec of silence at the beginning of the second talker's layer, mixing the layers and then trimming the first 500 msec. Consequently, the initial 500 msec of the first talker's first sentence was not included in the two-talker babble. For each noise condition, the target stimulus /da/ and the appropriate babble track were mixed and presented continuously in the background by Stim Sound Editor (Compumedics, Charlotte, NC) at a SNR of +10 dB. We examined the effect of different levels of energetic masking (two-talker vs. six-talker babble) on the subcortical encoding of the F_0 of the target speech sound while holding the SNR constant. The two-talker babble and the six-talker babble were composed of nonsense sentences that were looped for the duration of data collection (approximately 25 min per condition) with no silent intervals, such that the /da/ and the background babble noise were repeated at different noninteger time points. This presentation paradigm allowed the noise to have a randomized phase with respect to the target speech sound. Thus, responses that were time locked to the target sound could be averaged without the confound of having phase-coherent responses to the background noise.

During testing, the participants watched a captioned video of their choice with the sound level set at <40 dB SPL to facilitate a passive yet wakeful state. In each of the quiet and two noise conditions condition, 6300 sweeps of /da/ were presented. Responses were collected using Scan 4.3 Acquire (Compumedics) in continuous mode with Ag–AgCl scalp electrodes differentially recording from Cz (active) to right earlobe (reference), with the forehead as ground at a 20-kHz sampling rate. The continuous recordings were filtered, artifact rejected (±35 μV), and averaged off-line using Scan 4.3. Responses were band-pass filtered from 70 to 1000 Hz, 12 dB/octave. Waveforms were averaged with a time window spanning 40 msec before the onset and 16.5 msec after the offset of the stimulus and baseline corrected over the prestimulus interval (−40 to 0 msec, with 0 corresponding to the stimulus onset). Responses of alternating polarity were added to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Gorga, Abbas, & Worthington, 1985). The final average response consisted of the first 6000 artifact-free responses. For additional information regarding stimulus presentation and brainstem response collection parameters, refer to Skoe and Kraus (2010).

Neurophysiologic Analysis Procedures

Representation of Fundamental Frequency

ROIs. Within the FFR, the F_0 analysis was divided into two regions on the basis of the autocorrelation of the stimulus: (1) a transition region (20–60 msec) corresponding to the neural response to the transition within the stimulus as the syllable /da/ proceeds from the stop-consonant to the vowel and (2) a steady-state region (60–180 msec) corresponding with the encoding of the vowel. As can be seen in the autocorrelogram (a visual display of the sound periodicity; Figure 1C) of the stimulus, the temporal features of the stimulus are more variable in the time region corresponding to the formant transition. For example, at the period of the fundamental frequency (lag = 10 msec), the autocorrelation function has a considerably lower average r value during the formant transition compared with the steady-state region (r = .14 and .84, respectively). This indicates that the F_0 is less periodic and consequently more variable in its phase during the formant frequency region. In the steady-state region, both the F_0 and the formants are constant, and consequently the temporal features of the stimulus are more constant. In contrast, the rapidly changing formants interact with the F_0 to produce more fluctuating temporal cues at the periodicity of the F_0 in the stimulus. Because the temporal cues relating to F_0 are more variable during the formant transition, we predicted that phase locking would be more variable, especially in individuals who are more susceptible to noise-induced neural desynchronization.

The division of the response into these two sections was also motivated by (1) previous demonstrations that the F_0 and formant frequencies interact in the ABR (Hornickel et al., 2009; Johnson et al., 2008), (2) two recent studies

showing that SIN perception correlates with subcortical timing during the formant transition but not the steady-state region, and (3) evidence that rapidly changing formant transitions pose particular perceptual challenges (Assmann & Summerfield, 2004; Merzenich et al., 1996; Tallal & Piercy, 1974; Miller & Nicely, 1955). Thus, the analysis of the F_0 was performed separately on the transition and steady-state regions to assess the possible differences in the strength and accuracy of neural encoding in each portion of the response. Figure 2A shows the top and the bottom groups' grand average speech ABRs in the three listening conditions: quiet, two-talker babble, and six-talker babble. Individual responses were segmented into the two time ranges of 20–60 msec (see Figure 2B) and 60–180 msec (see Figure 2C).

Fast Fourier analysis. The strength of F_0 encoding in the transition and steady-state regions of the response elicited by the different listening conditions were exam-ined in the frequency domain using the fast Fourier transform. The strength of F_0 encoding was defined as the average spectral amplitude within a 40-Hz wide bin centered around the F_0 (80–120 Hz). To quantify the amount of degradation in the noise (relative to quiet), we computed amplitude ratios as $[F_0(\text{quiet}) - F_0(\text{noise})] / F_0(\text{quiet})$ for the two noise conditions.

RESULTS

SIN Perception

Subjects were grouped on the basis of their performance on the 0-dB SNR condition of the QuickSIN test. Subjects with a SIN performance score equal to or better than the group median were placed in the "top" SIN group ($n = 9$, mean = 40.56% correct words, $SD = 16.29\%$ correct words). The highest score that was obtained was 75%; thus, there was no ceiling effect at this SNR. Subjects with a SIN
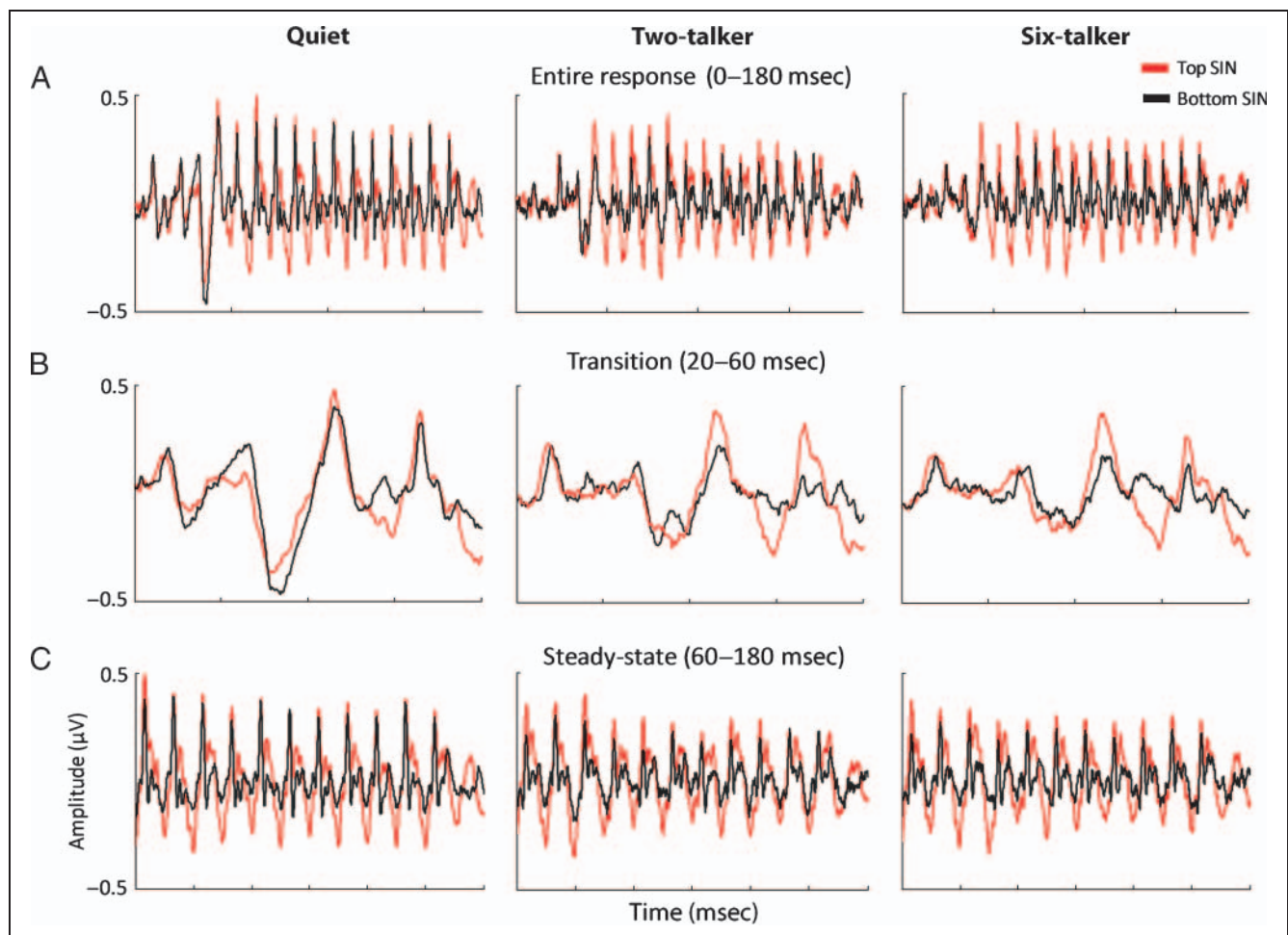
Figure 2. (A) Grand average brainstem responses of subjects with top (red) and bottom (black) SIN perception recorded to the /da/ stimulus without background noise (Quiet, left) and in two background noise conditions, two-talker (middle) and six-talker (right) babbles. (B) Overlay of top and bottom SIN groups' transition (20–60 msec) and (C) steady-state response (60–180 msec) show that the top SIN group has better representation of the F_0 in both background noise conditions as demonstrated by larger amplitudes of the prominent periodic peaks occurring every 10 msec. The transition portion of the response reflects the shift in formants as the stimulus moves from the onset burst to the vowel portion. The steady-state portion is a segment of the response that reflects phase locking to stimulus periodicity in the vowel.
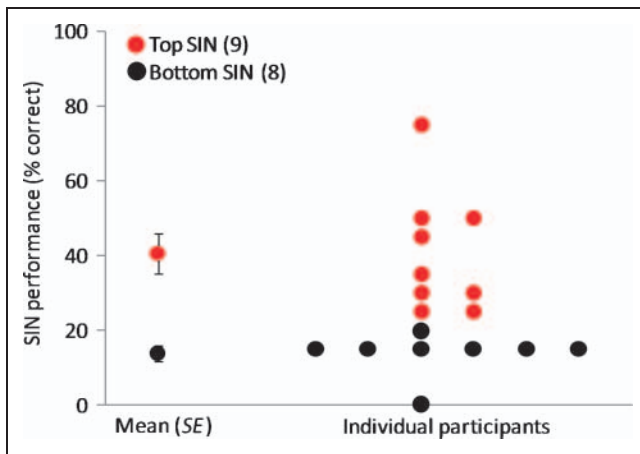
Figure 3. Average score (±1 *SE*) and distribution of individual subject's SIN performance (percent correct on QuickSIN). This measure was derived from the 0 dB SNR condition by dividing the number of correctly repeated target words from the final sentence of four randomly selected QuickSIN lists SNR. Subjects were categorized into top (≥25%, *n* = 9, red) and bottom (<25%, *n* = 8, black) SIN perceiving groups.

performance score below the median were grouped into a "bottom" SIN group (*n* = 8, mean = 13.75% correct words, *SD* = 5.83 correct words). Only one subject performed at floor, with 0 of 20 correct words. Figure 3 shows the distribution of SIN performance scores and the average score of each SIN group. The two groups significantly diverged in the 0-dB SNR condition (independent sample *t* test; *t* = −4.847, *p* = .001).

Brainstem Responses to Speech in Quiet and Noise

Formant Transition Period (20–60 msec): F_0 Representation

Background noise diminished the F_0 amplitudes in both listener groups but particularly in the bottom SIN group (Figure 4A). A 2 (group: top vs. bottom SIN) × 3 (F_0 amplitude of each listening condition: quiet, two-talker babble, and six-talker babble) repeated measures ANOVA, using the Greenhouse–Geisser correction to guard against violations of sphericity, revealed a significant main effect of listening condition (*F* = 21.325, *p* = .001). That is, the presentation of background noise significantly degraded the F_0 amplitude for all subjects. Although there was no main effect of group (*F* = 2.039, *p* = .196), the interaction between group and listening condition was significant (*F* = 6.181, *p* = .035). This result indicates that the presence of noise in the background degraded the F_0 amplitude in one group to a greater extent than the other group. Specifically, post hoc pairwise Bonferroni-corrected *t* tests showed that whereas the F_0 amplitudes of each SIN group were comparable in quiet (*p* = .99), the bottom SIN group's responses were significantly reduced in both noise conditions (*p* = .0151 and .0351, for two-talker and six-talker conditions, respectively). Furthermore, although

the F_0 amplitude of the top SIN group was not significantly reduced by the two-talker noise compared with quiet (*p* = .10), the F_0 amplitude was reduced for the bottom SIN group (*p* = .0029; see Figure 4B). However, both groups were significantly affected by the introduction of the six-talker noise (compared with quiet, *p* = .0039 and .0022 in the top and bottom groups, respectively). Effect sizes (using Cohen's *d*) of the group differences in noise (top vs. bottom SIN groups) were large for both noise conditions (*d* = 1.03 and 1.15 for the two-babble and six-babble conditions, respectively). Comparisons between listening conditions (i.e., quiet vs. two-talker condition, quiet vs. six-talker condition) within each group showed larger effect sizes in the bottom SIN group (within group Cohen's *d* = 2.73 and 2.83, respectively) compared with the top SIN group (*d* = 0.826 and 1.61, respectively). Thus, in quiet, there were no physiologic differences between top and bottom SIN perceivers, and the deterioration of F_0 encoding was significantly less in the top SIN group in both masking conditions.

Steady-state Period (60–180 msec): F_0 Representation

During the FFR period, F_0 amplitudes in the bottom SIN group tended to be smaller than those of the top SIN group (see Figure 4C and D). Although statistical differences between the groups failed to reach significance (repeated measures ANOVA on the F_0 in the steady-state period; group effect: *F* = 3.381, *p* = .109; condition effect: *F* = 1.197, *p* = .328; interaction effect: *F* = .214, *p* = .765), there was a trend for a SIN group difference in the talker conditions (two-talker condition: *t* = −2.28, *p* = .044; six-talker condition: *t* = −2.327, *p* = .042; quiet: *t* = −1.829, *p* = .09), and the effect sizes of the group differences were large in all three conditions (*d* = 0.99, 1.03, and 1.08 for quiet, two-talker, and six-talker noise conditions, respectively). Therefore, the F_0 amplitude during this period was smaller in the bottom SIN group irrespective of condition.

Relationship to Behavior

To examine the relationship between neural encoding of the F_0 and perception of SIN, we correlated the F_0 amplitudes obtained from the transition and steady-state periods with QuickSIN performance scores (α = .05). For the transition period of the response recorded in the six-babble condition, speech-evoked F_0 amplitude correlated positively with SIN performance (r_s = .523, *p* = .031) and approached significance in the two-talker babble condition (r_s = .459, *p* = .064; see Figure 5A). There was no significant relationship between F_0 response amplitude recorded in quiet and SIN performance (r_s = .009, *p* = .972). The degree of change in the F_0 amplitude from quiet to six-talker babble (larger values mean more degradation in noise) negatively correlated with SIN performance (r_s = −.593, *p* = .012), and these correlations approached significance in the two-talker babble condition (r_s = −.47, *p* = .057; see Figure 5B), indicating that the extent of

response degradation in noise relative to quiet contributes to SIN perception. These findings suggest that subcortical representation of the F_0 plays a role in the perception of speech in noisy conditions.

No significant correlations exist between audiometric thresholds (i.e., individual thresholds from 125 to 8k Hz; pure-tone average of 500, 1000, and 2000 Hz; and overall average of thresholds between 125 and 8000 Hz of each ear) and the F_0 amplitude or QuickSIN performance score.

DISCUSSION

Our results demonstrate that the strength of F_0 representation of speech is related to the accuracy of speech perception
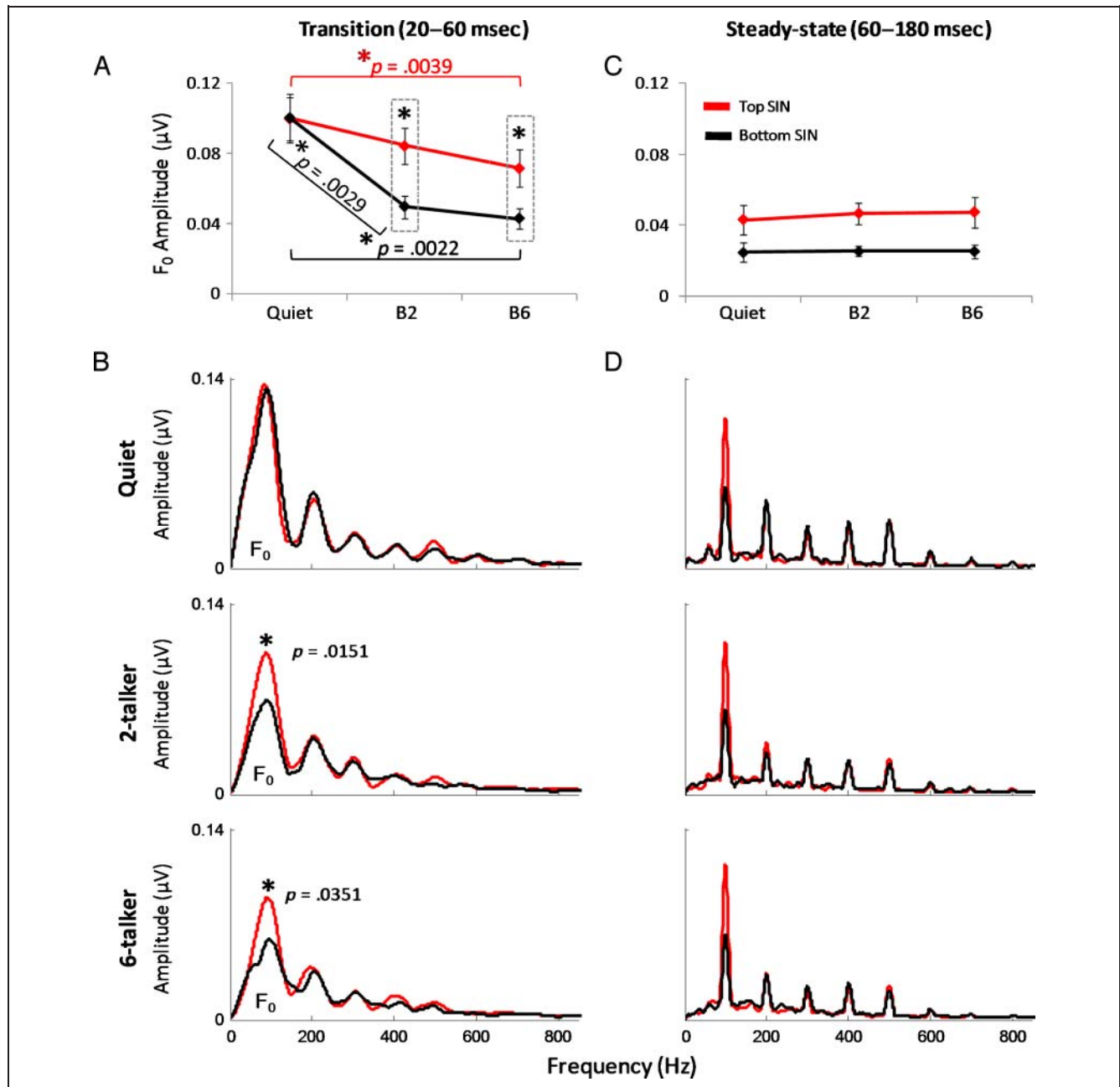
Figure 4. (A) Average fundamental frequency (F_0) amplitude (100 Hz) of the transition response (20–60 msec) for the top (red) and bottom (black) SIN groups for each listening condition (± 1 SE). (B) Grand average spectra of the transition response collected in quiet (top), two-talker (middle), and six-talker (bottom) noise for top and bottom SIN groups. For both noise conditions (B2 and B6), brainstem representation of the F_0 was degraded to a greater extent in the bottom SIN group relative to the top SIN group ($p = .0151$ and .0351, respectively). (C) Average F_0 amplitude of the steady-state portion (60–180 msec) for the top and bottom SIN groups for each listening condition (± 1 SE). The effect sizes of the group differences were large in all three conditions ($d = 0.99$, 1.03, and 1.08 for quiet, two-talker, and six-talker noise conditions, respectively). The top SIN group demonstrated stronger F_0 encoding in response to the sustained periodic vowel portion of the stimulus in all conditions. (D) Grand average spectra of the steady-state responses.

Figure 5. (A) Speech ABR F_0 amplitude of the formant transition period obtained from two-talker (left) and six-talker (right) babble conditions as a function of SIN performance for each subject. Magnitude of the F_0 correlated positively with SIN performance in the six-talker babble condition ($r_s = .523$, $p = .031$) and approached significance in the two-talker babble condition ($r_s = .459$, $p = .064$). (B) Normalized difference between quiet-to-noise F_0 amplitude for two-talker (left) and six-talker (right) conditions (i.e., $[F_0(quiet) - F_0(noise)] / F_0(quiet)$) as a function of SIN performance for each subject. Amplitude of the F_0 for both conditions related to SIN performance (two-talker $r_s = -.47$, $p = .057$ and six-talker $r_s = -.593$, $p = .012$). The dashed horizontal lines depict the linear fit of the F_0 amplitude and SIN measures.
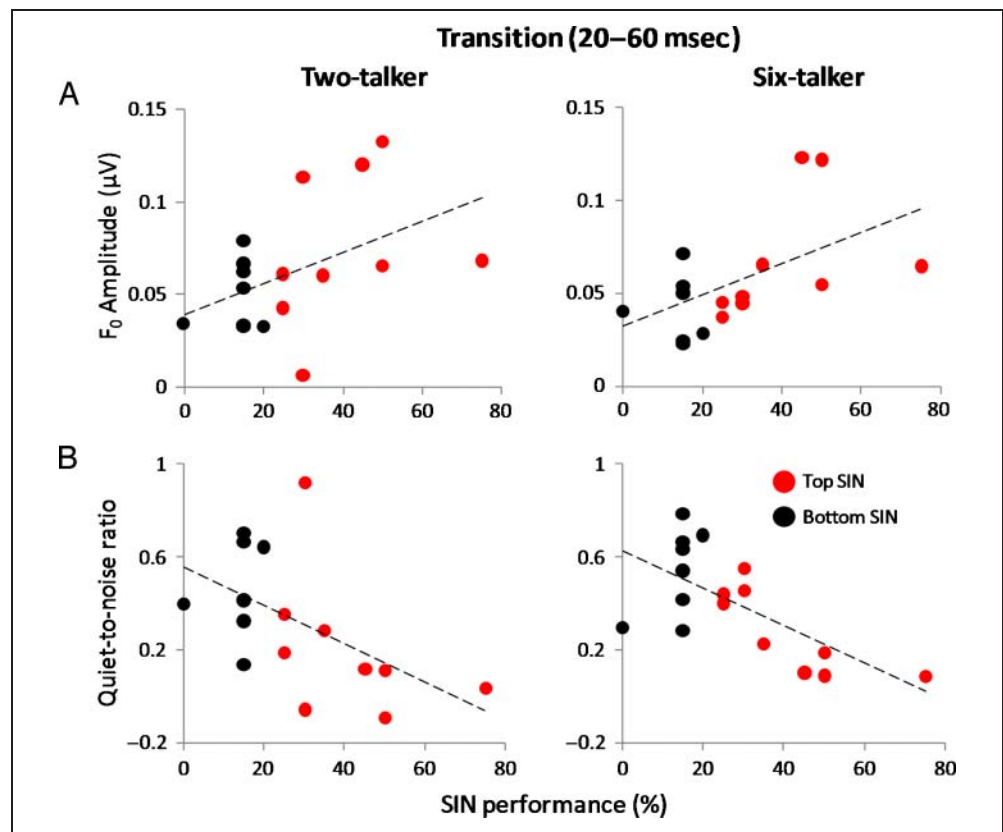
in noise. Listeners who exhibit poorer SIN perception are more susceptible to degradation of F_0 encoding in response to a speech sound presented in noise. These findings suggest that subcortical neural encoding could be one source of individual differences in SIN perception, thereby furthering our understanding of the biological processes involved in SIN perception.

The F_0 of speech sounds is one of several acoustic features (e.g., formants, fine structure) that contribute to speech perception in noise. The F_0 is a robust feature that offers a basis for grouping speech units across frequency and over time (Assmann & Summerfield, 2004; Darwin & Carlyon, 1995). It signals whether two speech sounds were produced by the same larynx and vocal tract (Langner, 1992; Assmann & Summerfield, 1990; Bregman, 1990), thus making it important for determining speaker identity. F_0 variation also underlies the prosodic structure of speech and helps listeners select among alternative interpretations of utterances especially when they are partially masked by other sounds (Assmann & Summerfield, 2004). When the target voice is masked by other voices, listeners find it easier to understand the message while tracking the F_0 of the desired speaker (Assmann & Summerfield, 2004; Bird & Darwin, 1998; Brokx & Nooteboom, 1982), and presumably this would affect the perception of elements riding on the F_0 (i.e., pitch and formants). The current data suggest that individuals who are less susceptible to the degradation of F_0 representation at the level of the brainstem due to background noise may be at an advantage when it

comes to tracking the F_0, aiding in their speech perception in noise.

We have shown an association between normal variation in SIN perception and brainstem encoding of the F_0 of speech presented in noise. This relationship between the strength of F_0 representation and perception of SIN is particularly salient in the portion of the syllable in which the periodicity of the F_0 is weakened by rapidly changing formants. Moreover, brainstem encoding of the F_0 in individuals with poorer SIN perception is affected to a greater extent by noise than those with better SIN perception. In fact, in the face of less spectrally dense noise (two-talker babble), the F_0 magnitude of the top SIN group's brainstem response did not differ significantly from their response in quiet, whereas the F_0 representation in the bottom SIN group was more susceptible to the deleterious effects of both noise conditions. Although both groups showed diminished F_0 representation in the most spectrally dense background noise condition (six-talker babble) relative to quiet, this reduction was greater for the bottom SIN group. Consequently, the perceptual problems associated with diminished speech discrimination in background noise may be attributed in part to the decreased neural synchrony that leads to decreased F_0 encoding.

Even in this selective sample of normal young adults, sufficient variability in both SIN perception and brainstem function was observed with no difference in speech perception in quiet. If this is the case among normal hearing young adults, we would expect differences to be more

pronounced in clinical populations where SIN perception is deficient. Further studies are required to elucidate neurophysiologic and cognitive processes involved in listeners who are generally "good speech encoders" (i.e., enhanced performers even in quiet) and those who have SIN perception problems. It should be noted that the use of monaural stimulation during the recording of brainstem responses limits the generalizability of our findings to more real-world listening conditions in which both ears are involved in perception. Additional studies are needed to determine the generalizability of these findings to brainstem responses to binaural stimulation.

Robust stimulus encoding in the auditory brainstem may affect cortical encoding in a feed-forward fashion by propagating a stronger neural signal, ultimately enhancing SIN performance. The relationship between perceptual and neurophysiologic processes can be also viewed within the framework of corticofugal (top–down) tuning of sensory function. It is known that modulation of the cochlea can facilitate speech perception in noise via the descending auditory pathway (deBoer & Thorton, 2008; Luo, Wang, Kashani, & Yan, 2008). Moreover, participants with better SIN perception may have learned to use cognitive resources to better attend to and integrate target speech cues and to use contextual information in the midst of background babble (Shinn-Cunningham, 2008; Shinn-Cunningham & Best, 2008). Thus, top–down neural control may enhance subcortical encoding of the F_0-related information of the stimuli. Cortical processes project backward to tune structures in the auditory periphery (Zhang & Suga, 2000, 2005); in the case of speech perception in noise, these processes may enhance features of the target speech sounds subcortically (Anderson, Skoe, Chandrasekaran, Zecker, & Kraus, in press). Such enhancement may allow the listener to extract pertinent speech information from background noise, consistent with perceptual learning models involving changes in the weighting of perceptual dimensions because of feedback (Amitay, 2009; Nosofsky, 1986). These models suggest an increase in weighting of parameters relevant to important acoustic cues (Kraus & Chandrasekaran, 2010) such as the F_0 when listening in noise, especially in those who exhibit better SIN perception. Life-long and training-associated changes in subcortical function are consistent with corticofugal shaping of subcortical sensory function (Kraus, Skoe, Parbery-Clark, & Ashley, 2009; Krishnan et al., 2009). The significant correlation between the six-talker condition and the SIN perception, and not for the quiet or two-talker conditions in our data, suggests that the brainstem representation of the F_0 (low-level information) may be exploited in situations that require a better SNR (i.e., six-talker imposing greater spectral masking) guided by top–down-activated pathways. This notion corresponds with the Reverse Hierarchy Theory, which states that more demanding task conditions require greater processing at lower levels (Ahissar & Hochstein, 2004).

The auditory cortex is unquestionably involved when listening to spoken words in noisy conditions (Scott, Rosen,

Beaman, Davis, & Wise, 2009; Bishop & Miller, 2009; Gutschalk, Micheyl, & Oxenham, 2008; Wong, Uppunda, Parrish, & Dhar, 2008; Obleser, Wise, Dresner, & Scott, 2007; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006; Scott, Rosen, Wickham, & Wise, 2004; Boatman, Vining, Freeman, & Carson, 2003; Martin, Kurtzberg, & Stapells, 1999; Shtyrov et al., 1998). Relative to listening to speech in quiet, listening in noise increases activation in a network of brain areas, including the auditory cortex, particularly the right superior temporal gyrus (Wong et al., 2008). The manner in which subcortical and cortical processes interact to drive experience-dependent cortical plasticity remains to be determined (Bajo, Nodal, Moore, & King, 2010). Nevertheless, brainstem-cortical relationships in the encoding of complex sounds have been established in humans (Abrams, Nicol, Zecker, & Kraus, 2006; Banai, Nicol, Zecker, & Kraus, 2005; Wible, Nicol, & Kraus, 2005) and in particular for SIN (Wible et al., 2005). Most likely, a reciprocally interactive, positive feedback process involving sensory and cognitive processes underlies listening success in noise.

Speech-evoked responses provide objective information about the neural encoding of speech sounds in quiet and noisy listening conditions. These brainstem responses also reveal subcortical processes underlying SIN perception, a task that depends on cognitive resources for the interpretation of limited and distorted signals. Better understanding of how noise impacts brainstem encoding of speech in young adults, in this case the strength of F_0 encoding, serves as a basis for future studies investigating the neural mechanisms underlying SIN perception. From a clinical perspective, our findings may provide an objective measure to monitor training-related changes and to assess clinical populations with excessive difficulty hearing SIN such as individuals with language impairment (poor readers, SLI, APD), hearing impairment, older adults, and nonnative speakers. The establishment of a relationship between neural encoding and perception of an important speech cue in noise (i.e., F_0) is a step toward this goal.

Acknowledgments

REFERENCES

Abrams, D. A., Nicol, T., Zecker, S. G., & Kraus, N. (2006). Auditory brainstem timing predicts cerebral asymmetry for speech. *Journal of Neuroscience, 26,* 11131–11137.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8,* 457–464.

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245,* 35–47.

Akhoun, I., Gallégo, S., Moulin, A., Menard, M., Veuillet, E., Berger-Vachon, C., et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology, 119,* 922–933.

Amitay, S. (2009). Forward and reverse hierarchies in auditory perceptual learning. *Learning & Perception, 1,* 59–68.

Ananthanarayan, A. K., & Durrant, J. D. (1992). The frequency-following response and the onset response: Evaluation of frequency specificity using a forward-masking paradigm. *Ear and Hearing, 13,* 228–232.

Anderson, S., Chandrasekaran, B., Skoe, E., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *Journal of Neuroscience, 30,* 4922–4926.

Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S., & Kraus, N. (in press). Brainstem correlates of speech-in-noise perception in children. *Hearing Research*.

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

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels—Vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 88,* 680–697.

Bajo, V. M., Nodal, F. R., Moore, D. R., & King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature Neuroscience, 13,* 253–256.

Banai, K., Nicol, T., Zecker, S. G., & Kraus, N. (2005). Brainstem timing: Implications for cortical processing and literacy. *Journal of Neuroscience, 25,* 9850–9857.

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research, 74,* 110–120.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23,* 425–434.

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

Bishop, C. W., & Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *Journal of Cognitive Neuroscience, 21,* 1790–1804.

Boatman, D., Vining, E. P., Freeman, J., & Carson, B. (2003). Auditory processing studied prospectively in two hemidecorticectonty patients. *Journal of Child Neurology, 18,* 228–232.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23–36.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence: A language-free measure of cognitive ability*. Austin, TX: PRO-ED Inc.

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron, 64,* 311–319.

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

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology, 112,* 758–767.

Darwin, C. J., & Carlyon, R. P. (1995). *Auditory grouping*. London: Academic Press.

deBoer, J., & Thorton, A. R. D. (2008). Neural correlates of perceptual learning in the auditory brainstem: Efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *Journal of Neuroscience, 28,* 4929–4937.

Galbraith, G. C., Amaya, E. M., de Rivera, J. M., Donan, N. M., Duong, M. T., Hsu, J. N., et al. (2004). Brain stem evoked response to forward and reversed speech in humans. *NeuroReport, 15,* 2057–2060.

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. (1995). Intelligible speech encoded in the human brain stem frequency following response. *NeuroReport, 6,* 2363–2367.

Galbraith, G. C., Threadgill, M. R., Hemsley, J., Salour, K., Songdej, N., Ton, J., et al. (2000). Putative measure of peripheral and brainstem frequency-following in humans. *Neuroscience Letters, 292,* 123–127.

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

Gutschalk, A., Micheyl, C., & Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biology, 6,* 1156–1165.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Subcortical differentiation of voiced stop consonants: Relationships to reading and speech in noise perception. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 13022–13027.

Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., & Kraus, N. (2008). Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology, 119,* 2623–2635.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & 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. *Journal of the Acoustical Society of America, 116,* 2395–2405.

King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters, 319,* 111–115.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11,* 599–605.

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical "what" and "where" pathways in the auditory system. *Trends in Neurosciences, 28,* 176–181.

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre and timing: Implications for language and music. *Annals of the New York Academy of Sciences, 1169,* 543–557.

Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research, 166,* 192–201.

Krishnan, A., Gandour, J., & Bidelman, G. (2010). Brainstem pitch representation in native speakers of Mandarin is

less susceptible to degradation of stimulus temporal regularity. *Brain Research, 1313,* 124–133.

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience, 21,* 1092–1105.

Krishnan, A., Xu, Y. S., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25,* 161–168.

Krishnan, A., Xu, Y. S., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: Representation of pitch contours in Chinese tones. *Hearing Research, 189,* 1–12.

Langner, G. (1992). Periodicity coding in the auditory-system. *Hearing Research, 60,* 115–142.

Lee, K., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience, 29,* 5832–5840.

Luo, F., Wang, Q. Z., Kashani, A., & Yan, J. (2008). Corticofugal modulation of initial sound processing in the brain. *Journal of Neuroscience, 28,* 11615–11621.

Marler, J. A., & Champlin, C. A. (2005). Sensory processing of backward-masking signals in children with language-learning impairment as assessed with the auditory brainstem response. *Journal of Speech, Language, and Hearing Research, 48,* 189–203.

Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. *Journal of Speech, Language, and Hearing Research, 42,* 271–286.

McAnally, K. I., & Stein, J. F. (1997). Scalp potentials evoked by amplitude-modulated tones in dyslexia. *Journal of Speech, Language, and Hearing Research, 40,* 939–945.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271,* 77–81.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338–352.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 15894–15898.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics, 41,* 409–415.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95,* 1085–1099.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–61.

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience, 27,* 2283–2289.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience, 29,* 14100–14107.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 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. *Developmental Science, 12,* 557–567.

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research, 156,* 95–103.

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., et al. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clinical Neurophysiology, 119,* 1720–1731.

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., & Wise, R. J. S. (2009). The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. *Journal of the Acoustical Society of America, 125,* 1737–1743.

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of the Acoustical Society of America, 115,* 813–821.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12,* 182–186.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification, 12,* 283–299.

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., et al. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: Magnetic mismatch negativity study. *Neuroscience Letters, 251,* 141–144.

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear and Hearing, 31,* 302–324.

Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America, 118,* 1677–1688.

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience, 20,* 1892–1902.

Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F. G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *Journal of the Acoustical Society of America, 122,* 1069–1078.

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency—Effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience, 29,* 661–668.

Summers, V., & Leek, M. R. (1998). F_0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research, 41,* 1294–1306.

Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport, 19,* 1163–1167.

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 [Computer software].

Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron, 62,* 463–469.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America, 121,* 519–526.

Wang, J., Nicol, T., Skoe, E., Sams, M., & Kraus, N. (2009). Emotion modulates early auditory response to speech. *Journal of Cognitive Neuroscience, 21,* 2121–2128.

Wible, B., Nicol, T., & Kraus, N. (2005). Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain, 128,* 417–423.

Wilson, J. R., & Krishnan, A. (2005). Human frequency-following responses to binaural masking level difference stimuli. *Journal of the American Academy of Audiology, 16,* 184–195.

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 50,* 844–856.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10,* 420–422.

Wong, P. C. M., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *Journal of Speech, Language, and Hearing Research, 51,* 1026–1041.

Xu, Y. S., Krishnan, A., & Gandour, J. T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *NeuroReport, 17,* 1601–1605.

Yamada, O., Kodera, K., Hink, F. R., & Suzuki, J. I. (1979). Cochlear distribution of frequency-following response initiation. A high-pass masking noise study. *Audiology, 18,* 381–387.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *Neuroimage, 32,* 1826–1836.

Zhang, Y. F., & Suga, N. (2000). Modulation of responses and frequency tuning of thalamic and collicular neurons by cortical activation in mustached bats. *Journal of Neurophysiology, 84,* 325–333.

Zhang, Y. K., & Suga, N. (2005). Corticofugal feedback for collicular plasticity evoked by electric stimulation of the inferior colliculus. *Journal of Neurophysiology, 94,* 2676–2682.