



Research paper

Brainstem correlates of speech-in-noise perception in children

Samira Anderson^{a,*}, Erika Skoe^a, Bharath Chandrasekaran^b, Steven Zecker^a, Nina Kraus^{a,c,d}^a Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA^b Communication Sciences and Disorders, 1 University Station A1100, The University of Texas at Austin, TX 78712, USA^c Department of Neurobiology and Physiology, Northwestern University, 2205 Tech Drive, Evanston, IL 60208, USA^d Department of Otolaryngology, Northwestern University Feinberg School of Medicine, 303 E. Chicago Ave., Chicago, IL 60611, USA

ARTICLE INFO

Article history:

Received 1 June 2010

Received in revised form

30 July 2010

Accepted 4 August 2010

Available online 12 August 2010

ABSTRACT

Children often have difficulty understanding speech in challenging listening environments. In the absence of peripheral hearing loss, these speech perception difficulties may arise from dysfunction at more central levels in the auditory system, including subcortical structures. We examined brainstem encoding of pitch in a speech syllable in 38 school-age children. In children with poor speech-in-noise perception, we find impaired encoding of the fundamental frequency and the second harmonic, two important cues for pitch perception. Pitch, an essential factor in speaker identification, aids the listener in tracking a specific voice from a background of voices. These results suggest that the robustness of subcortical neural encoding of pitch features in time-varying signals is a key factor in determining success with perceiving speech in noise.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In our industrial society, increased noise levels can interfere with communication and learning. Noise levels and reverberation times in today's classrooms commonly exceed recommended levels (Summers and Leek, 1998; ANSI, 2002; Knecht et al., 2002; Krishnan and Gandour, 2009), hindering communication and putting children at risk for academic failure. In fact, classroom noise levels are directly related to scholastic achievement (Shield and Dockrell, 2003, 2008), and increases in noise of as little as 10 dB can result in a significant decrease in academic test scores. Some children seem to have more trouble adapting to background noise than others, and many explanations for these difficulties have been proposed, including the presence of auditory processing disorders (Moore, 2007; Lagace et al., 2010), language-based learning disorders (Bradlow et al., 2003; Ziegler et al., 2005, 2009), attention deficit disorders (Abdo et al., 2010), or noise exclusions deficits (Sperling et al., 2005). In the absence of peripheral hearing loss, these speech perception difficulties may arise from dysfunction at more central levels in the auditory system, including subcortical structures. In the current study, we explored the role of subcortical function in speech-in-

noise (SIN) perception by measuring brainstem responses to a speech syllable in a group of children with a broad range of SIN abilities. We expected to see group differences between top and bottom SIN perceivers in their encoding of the acoustic cues linked to SIN perception, particularly those involving pitch.

Brainstem testing is ideally suited for pediatric evaluation, as it is objective, noninvasive and pre-attentive (Hall and Mueller, 1997). The auditory brainstem response to speech (speech-ABR) is highly replicable (Kraus and Nicol, 2005; Song et al., in press), with latency differences on the order of fractions of milliseconds being clinically significant (Wible et al., 2004). The brainstem's representation of the stimulus is remarkably robust; in addition to the brainstem response waveform being visually similar to the stimulus waveform (Kraus et al., 2009; Krishnan and Gandour, 2009), the audio file of the brainstem response is also acoustically similar to the stimulus (Galbraith et al., 1995; Skoe and Kraus, 2010). Importantly, brainstem encoding of speech features is known to be experience-dependent (Krishnan et al., 2005; Song et al., 2008; Tzounopoulos and Kraus, 2009) and relates to cortical (Banai et al., 2005; Wible et al., 2005; Musacchia et al., 2009) and cognitive processes such as language (Krishnan et al., 2005; Banai et al., 2009; Krishnan and Gandour, 2009) and music experience (reviewed in Kraus and Chandrasekaran, 2010).

The brainstem response to speech syllables can be used as a neural index of asynchrony in impaired populations. For example, children with learning impairments (Cunningham et al., 2001; Wible et al., 2004;) and reading disabilities (Banai et al., 2005, 2009; Chandrasekaran et al., 2009a; Hornickel et al., 2009; Anderson

Abbreviations: speech-in-noise, SIN; fundamental frequency, F_0 ; second harmonic, H_2 ; Hearing in Noise Test, HINT; fast Fourier transform, FFT; auditory brainstem response, ABR; signal-to-noise ratio, SNR; root mean square, RMS.

* Corresponding author. Tel.: +1 847 491 2465.

E-mail address: sba@u.northwestern.edu (S. Anderson).

URL: <http://www.brainvolts.northwestern.edu>

et al., 2010) exhibit neural asynchrony, or temporal jitter, leading to delayed peak timing and/or broadened peaks compared to controls. These response differences are especially apparent in the response to the syllable onset and to the consonant–vowel formant transition.

Given the demonstrated utility of the brainstem response in measuring neural asynchrony in children with language-based learning impairments and the findings that SIN perception relates to brainstem encoding of stop consonants (Hornickel et al., 2009; Anderson et al., 2010), we hypothesized that the speech-ABR may provide a biological marker for children with impaired SIN perception. Although previous studies have not demonstrated a relationship between fundamental frequency (F_0) encoding and reading (Banai et al., 2009), the strength of the F_0 plays an important role in SIN perception in young adults (Song et al., in press). We hypothesized that poor SIN perception results in part from impaired neural mechanisms responsible for pitch encoding and that we would therefore see neural correlates of degraded pitch representation in the brainstem responses of normal-hearing children with poor SIN perception.

Many acoustic ingredients contribute to the perception of pitch (Fellowes et al., 1997). In particular, the F_0 and the lower harmonics provide essential acoustic cues relating to pitch (Meddis and O'Mard, 1997), and these cues, along with spatial, timing, harmonic, and other cues, aid in speaker identification and object formation, two processes that are integral for SIN perception (Oxenham, 2008; Shinn-Cunningham and Best, 2008). Object formation (grouping of spectrotemporal cues to form a perceptual identity) is a necessary step in stream segregation, a process that allows the listener to extract meaning from an auditory scene consisting of simultaneous inputs (Bregman, 1990). When listening to one voice in a background of voices and other noises, the listener must perceive that voice as a distinct auditory stream. According to Oxenham (2008), auditory streaming involves two stages: simultaneous grouping (perceiving simultaneously-presented sounds as separate entities) and sequential grouping (binding together of elements that belong to the same sound source). The F_0 directly influences both sequential and simultaneous grouping. For example, when listeners are presented with two vowels played simultaneously, vowel identification improves as the F_0 differences increase (Brokx and Nootboom, 1982; Scheffers, 1983; Assmann and Summerfield, 1987; Culling and Darwin, 1993; de Cheveigne, 1997), and the extent to which F_0 separation aids identification is related to F_0 discrimination thresholds in normal-hearing and hearing-impaired listeners (Summers and Leek, 1998). Similarly, when listening to competing sentences, increasing the F_0 separation between concurrent sentences improves performance (Brokx and Nootboom, 1982; Bird and Darwin, 1998). This behavioral advantage with greater separation in the F_0 may reflect neurobiological processes at both cortical (Alain et al., 2005) and subcortical levels (Song et al., in press; Anderson and Kraus, in press). The importance of the lower harmonics in pitch perception has also been demonstrated in behavioral (Moore et al., 1985, 2005; Lin and Hartmann, 1998) and electrophysiological studies (Alain et al., 2001; Krishnan et al., 2005). Thus, impairments in object formation and pitch representation may be a factor in the SIN perception deficits experienced by clinical populations such as children with language-based learning disabilities and hearing-impaired listeners.

2. Materials and methods

2.1. Participants

Participants were recruited from public and private schools in the Chicago area as part of an ongoing study examining neural encoding of speech in children. Thirty-eight typically developing

children (ages 8 to 14; 22 male) with a wide range of SIN perception abilities participated. All children underwent otoscopic examination and were excluded if they had tympanic membrane perforation, cerumen occlusion, or middle ear effusion. Audiometric thresholds were measured at octave intervals from 250 to 8000 Hz using air conduction and from 500 to 4000 Hz using bone conduction. Participants were excluded if they had pure-tone thresholds greater than 15 dB in either ear or air-bone gaps ≥ 15 dB at two or more frequencies in either ear. Inclusionary criteria included normal Wave V click-evoked ABR latencies and normal cognitive abilities as evidenced by standard scores of ≥ 85 on the Wechsler Abbreviated Scales of Intelligence (WASI) Verbal, Performance and Full Scale scores (Zhu and Garcia, 1999). Based on these criteria, no children were excluded from participating in this study. The children were paid for their participation. Northwestern University's institutional review board approved all experimental procedures.

2.2. Behavioral measures

Speech understanding in noise was evaluated with a well-validated, commonly-used clinical instrument, the Hearing in Noise Test (HINT – Bio-logic Systems Corp., Mundelein, IL). HINT uses the Bamford-Kowal-Bench (BKB) phonetically-balanced sentences (Bench et al., 1979), which are appropriate for use with children at the first-grade reading level and above, and age-corrected percentile scores are available. HINT uses an adaptive paradigm that varies the intensity of the target sentence relative to the fixed speech-shaped noise masker until a threshold signal-to-noise ratio (SNR) is determined. Participants are told to ignore the noise and repeat lists of 10 BKB sentences. The experimenter counts a sentence as correct only when all words are repeated correctly. The full HINT protocol includes three conditions: HINT-Front (target sentences and noise are presented from a single loud speaker placed directly in front of the child), HINT-Right (target sentences are presented from the loud speaker in front of the child, and noise is presented from a loud speaker placed 90° to the right of the child), and HINT-Left (same as HINT-Right, except that the noise is presented from a loud speaker placed 90° to the left of the child). We chose to use the HINT-Front subtest in our analysis in order to eliminate the effect of spatial cues in the HINT-Right and HINT-Left conditions, thus forcing the listeners to rely only on acoustic cues such as pitch. Because previous studies have demonstrated a relationship between working memory and SIN performance (Pichora-Fuller and Souza, 2003; Heinrich et al., 2007; Parbery-Clark et al., 2009b), we assessed working memory using the Woodcock Johnson III Numbers Reversed subtest (Woodcock et al., 2001) to determine whether memory was a factor in our results.

2.3. Participant groups

The top SIN group included children with HINT-Front SNRs ≤ -1.0 ($N = 18$) and the bottom SIN group included children with HINT SNRs > -1.0 ($N = 20$). There were no significant differences between the top and bottom SIN groups for pure-tone average audiometric thresholds (0.5 kHz–4 kHz) ($t = -0.500$, $p = 0.621$), click-evoked ABR Wave V latencies ($t = -0.812$, $p = 0.428$), WASI full scale IQ standard scores ($t = 1.272$, $p = 0.212$), and Digits Reversed ($t = -0.471$, $p = 0.640$). See Table 1 for means and standard deviations for each measure.

2.4. Electrophysiology

2.4.1. Stimuli

The speech syllable [da] is a six-formant syllable synthesized at a 20 kHz sampling rate using a Klatt synthesizer (Klatt, 1980). The

Table 1

The means and standard deviations for top and bottom SIN groups are listed for HINT-Front scores, click latencies, WASI standard scores, WJIII Numbers Reversed standard scores, and pure-tone averages (0.5e4 kHz).

	HINT dB (SNR)	Click	WASI	Numbers	PTA for
		Latency Wave V (ms)	Full (Standard Score)	Rev. (Standard Score)	(.5–4kHz) dB HL
Total Group	–1.11(1.42)	5.89 (.11)	125.21 (13.34)	111.11 (17.82)	1.85 (3.53)
Mean (SD)					
Top SIN	–2.23 (1.19)	5.86 (.07)	127.80 (13.79)	109.80 (22.10)	1.45 (2.37)
Mean (SD)					
Bottom SIN	–.09 (.63)	5.90 (.12)	122.33 (12.60)	112.55 (11.87)	2.13 (4.21)
Mean (SD)					

duration of the syllable is 170 ms, the first five ms of which is a noiseburst, representing the initial unvoiced portion of the stop consonant preceding voicing. The formant transition from the [d] to the [a] is 50 ms in duration and is characterized by the following formant frequency trajectories: linearly rising F_1 (400–720 Hz) and linearly falling F_2 (1700–1240 Hz) and F_3 (2580–2500 Hz). F_4 (3300 Hz), F_5 (3750 Hz), and F_6 (4900 Hz) remain constant for the duration of the stimulus. The steady-state region (60 to 170 ms) has a constant F_0 of 100 Hz as well as a constant F_1 (720 Hz), F_2 (1240 Hz), F_3 (2500 Hz), F_4 (3300 Hz), F_5 (3750 Hz), and F_6 (4900 Hz).

2.4.2. Recording

Using NeuroScan Acquire 4.3 (NeuroScan Compumedics Inc., Charlotte, NC), brainstem responses were differentially recorded from Cz to right earlobe with forehead as ground at a sampling rate of 20 kHz. Stimuli were presented in the right ear through an insert earphone via an electromagnetically-shielded transducer (ER-3, Etymotic Research, Elk Grove Village, IL) using the stimulus presentation software NeuroScan Stim2. The left ear was unoccluded enabling the participants to hear movies or cartoons of their choice with the sound presented at <40 dB SPL. The use of movies ensured participant cooperation and enabled them to sit quietly for 2-h sessions.

The [da] syllable was presented at 80 dB SPL with a 60 ms inter-stimulus interval using interleaved alternating polarities. Because the two stimulus polarities are 180° out of phase, they can be added together to cancel out the cochlear microphonic and stimulus artifact (Gorga et al., 1985; Russo et al., 2004). Our approach emphasizes the envelope following response (Aiken and Picton, 2008; Skoe and Kraus, 2010), which does not invert between the two stimulus polarities.

2.5. Data analysis

Following data collection, the continuous response was filtered from 70 to 2000 Hz (12dB/octave, zero phase-shift) and epoched over a window of –40 to 190 ms with time 0 corresponding to the stimulus onset. This bandpass range was chosen to minimize low-frequency myogenic noise and cortical activity and to include energy that would be expected in the brainstem response given its phase locking properties (Chandrasekaran and Kraus, 2010b). An artifact reject criterion of $\pm 35 \mu\text{V}$ was applied, and for each stimulus polarity 3000 artifact-free responses were averaged together. The final average was generated by adding the averaged responses to the two polarities. To ensure response replicability, two sub-averages, representing the first half and last half of the recordings, were created for comparison purposes. The

two sub-averages and final added average waveforms were overlaid, and the response was considered reliable if the three waveforms were visually aligned and the first and last halves of the recording were replicable as quantified by a correlation coefficient of at least 0.55.

The SNR of the final average response, which was assessed by dividing the root mean square (RMS) of the response portion (0–190 ms) of the waveform by the RMS of the pre-stimulus portion (–40 to 0 ms), was used as a criterion for determining if the response was adequately free of extraneous myogenic and electrical noise. All subjects had a minimum SNR of 1.35 (mean: 2.31, S.D.: 0.54).

2.6. Measurement of the brainstem response

The brainstem evoked response to this 170-m [da] syllable is characterized by 2 general components: the onset response and the frequency following response (FFR) to the consonant-vowel transition and vowel. Neural phase locking that underlies the FFR represents the periodicity of the stimulus up to about 1500 Hz, the phase locking limit of the human brainstem (Moushegian et al., 1973; Chandrasekaran and Kraus, 2010b). The onset response occurs within 8–11 ms after the sound begins and is analogous to Wave V in the click-evoked ABR (Song et al., 2006;). The FFR includes the transition response and the steady-state response. The transition response occurs within the range of 20–60 ms and corresponds to the consonant-vowel transition in the stimulus. The steady-state response, from 60 to 170 ms, corresponds to the vowel [a]. The transition and the steady state are characterized by large, periodic peaks occurring approximately every 10 ms, corresponding to the periodicity of the syllable's F_0 of 100 Hz. The morphology of the transition peaks becomes sharper and more defined approaching 60 ms, the point at which the formants are constant, and the response is also less variable. The amplitudes of the peaks in the transition are modulated by changing formant frequencies (Johnson et al., 2008), while the peaks during the steady state are relatively stable in amplitude. Fig. 1, middle, depicts the major regions of the response.

The spectral energy in the frequency domain was analyzed in MATLAB (The MathWorks, Inc., Natick, MA) by computing fast Fourier transforms (FFTs) with zero padding to increase spectral resolution (Skoe and Kraus, 2010). Average spectral amplitudes were calculated for the formant transition period and the steady-state period using 100 Hz bins centered around the F_0 (100 Hz) and its harmonics (integer multiples of 100 Hz up to 1000 Hz). The resulting spectral magnitudes provided information regarding the representation of F_0 and its harmonics in the brainstem response as seen in Fig. 1, bottom.

2.7. Statistical analyses

Multivariate analyses of variance (MANOVA) were conducted using SPSS (SPSS Inc., Chicago, IL) with group (top SIN vs. bottom SIN) serving as the between-group independent variable. Consistent with our hypotheses, we expected to see significant differences in the low frequencies between the SIN groups. We therefore conducted 2 separate MANOVAs, one using the harmonic magnitudes associated with pitch (F_0 and H_2) and one using harmonics not typically associated with pitch (H_3 – H_{10}). Post-hoc *t*-tests were employed where appropriate. To assess the relationships between behavior and brainstem function, Pearson *r* correlations were calculated for the entire group between F_0 and H_2 and the HINT-Front SNR scores.

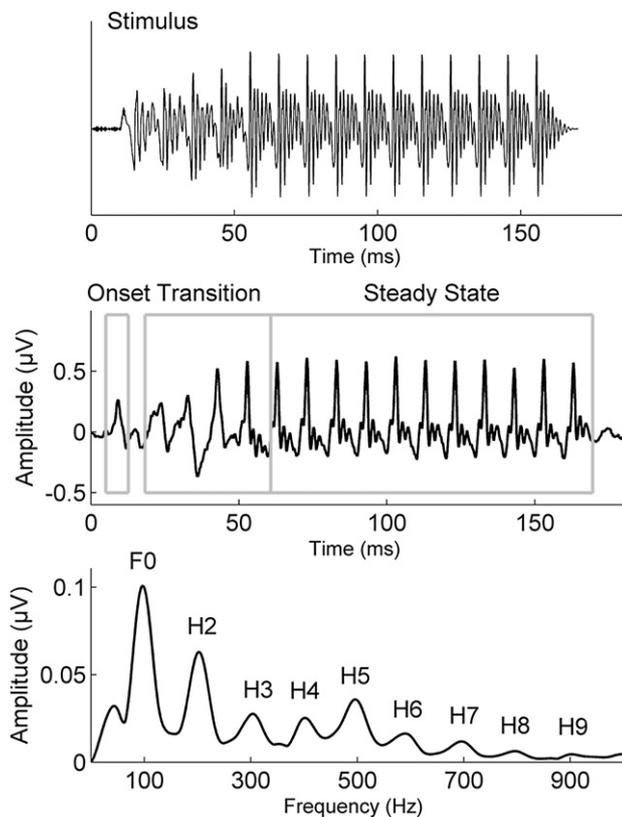


Fig. 1. Top: The stimulus waveform of the speech syllable [da]. Middle: Grand average brainstem responses ($N = 38$) to the speech syllable [da]. Bottom: The spectrum of the brainstem response to the speech syllable [da] contains energy at the fundamental frequency ($F_0 = 100$ Hz) and integer multiples (harmonics).

3. Results

3.1. Transition region

We found significant differences in low-frequency spectral encoding between top and bottom SIN perceivers in the region of the response corresponding to the formant transition period. There was a main effect of group in the frequency regions associated with pitch, with the top SIN group having greater spectral magnitudes for the F_0 and H_2 compared to the bottom SIN group ($F(2,35) = 3.926$, $p = 0.029$), and post-hoc t -tests revealed significant SIN group differences for the F_0 ($F(1,36) = 8.057$, $p = 0.007$) but not for H_2 ($F(1,36) = 0.547$, $p = 0.464$) (Fig. 2). There was no main effect of group when we included the higher harmonics (H_3 – H_{10}) as dependent variables in the MANOVA ($F(10,27) = 1.727$, $p = 0.126$).

3.2. Steady state region

High and low-performing SIN groups did not differ in their brainstem responses to the steady-state portion of the stimulus either for F_0 and H_2 ($F(1,36) = 1.298$, $p = 0.286$) or the higher harmonics ($F(10,27) = 1.232$, $p = 0.316$).

There were no significant differences in RMS ($t = -0.846$, $p = 0.403$) or SNR ($t = 0.798$, $p = 0.430$) between the groups. Therefore, the group differences in F_0 and H_2 encoding did not result from differences in overall neural activity.

3.3. Correlations

Pearson r correlations using the entire group ($N = 38$) were significant between HINT-Front and the spectral magnitudes of F_0

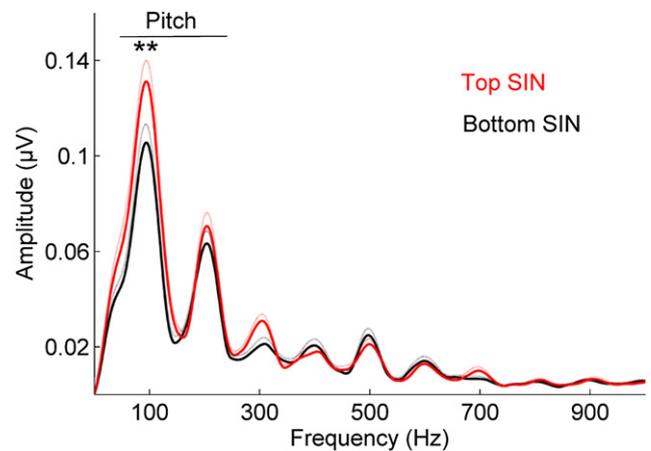


Fig. 2. Comparison of response spectra in top (red) and bottom (black) SIN perceivers. Group differences are found in the pitch-related lower harmonics ($F(2,35) = 3.926$, $p = 0.029$). Standard errors are plotted with dotted lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

($r = -0.424$, $p = 0.008$) and H_2 ($r = -0.336$, $p = 0.039$) in the response period corresponding to the formant transition, indicating that better performance on SIN tasks is associated with greater amplitudes of the frequency components contributing to pitch (Fig. 3). Because previous studies have demonstrated a relationship between working memory and SIN performance (Pichora-Fuller and Souza, 2003; Heinrich et al., 2007; Parbery-Clark et al., 2009b), we assessed working memory using the Woodcock Johnson III Numbers Reversed subtest to determine whether memory was a factor in our results. We found no relationship with performance on HINT ($r = -0.114$, $p = 0.495$). Furthermore, we found that working memory did not drive the group differences between HINT and pitch-related brainstem measures; all relationships remained significant when partialling out the effect of working memory (F_0 : $r = -0.430$, $p = 0.008$; H_2 : $r = -0.334$, $p = 0.044$).

4. Discussion

The novel finding from the current study is that the strength of subcortical encoding of pitch has emerged as a factor for successful hearing in noise in children. Pitch (determined by the low harmonics) has been previously identified as an important component for object formation and speech-in-noise perception in behavioral studies; the present study provides a neural mechanism for this observation. This neural mechanism operated exclusively in response to the time-varying formant transition period, a speech feature that is most vulnerable to masking by background noise (Tallal and Piercy, 1974; Tallal and Stark, 1981; Basu et al., 2010).

As noted previously, good readers have greater representation of formant-related higher harmonics (e.g. H_4 – H_7) in brainstem responses compared to poor readers, but the lower harmonics have not been implicated in poor readers. The differing importance of pitch-related lower harmonics for SIN perception and formant-related higher harmonics for reading suggests the existence of distinct neural signatures for these tasks.

Pitch cues may serve as perceptual anchors formed in response to stimulus regularities, and perceptual anchors enable the comparative discriminations needed for SIN perception (Ahissar et al., 2006; Ahissar, 2007; Chandrasekaran et al., 2009a). Ahissar et al. (2006) compared normal-learning children to children who had dyslexia combined with oral language disorders and found that the children with dyslexia had poorer SIN performance only when using small sets of stimuli but not when using large sets. They

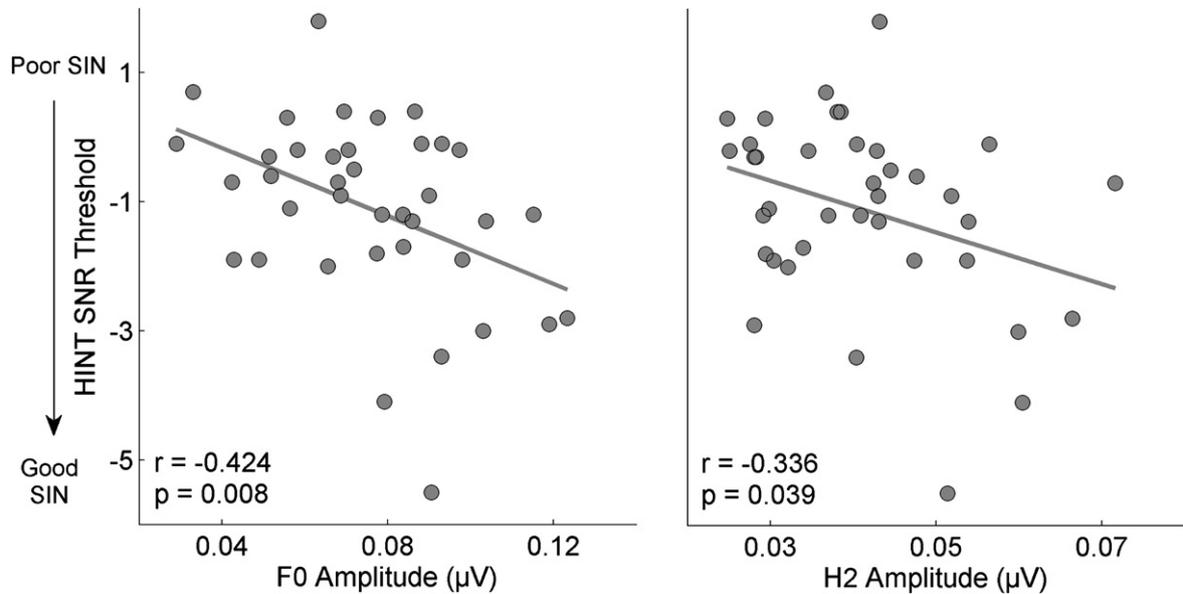


Fig. 3. HINT SNR thresholds are correlated with magnitudes of F_0 ($r = -0.424$, $p = 0.008$) and H_2 ($r = -0.336$, $p = 0.039$), indicating a relationship between SIN perception and subcortical encoding of pitch.

reasoned that the small set of stimuli allowed normal-learning children to form perceptual anchors, thereby profiting from repetition of the stimuli in order to improve performance. This context-dependent enhancement was recently noted in children using speech-evoked brainstem responses (Chandrasekaran et al., 2009a), in which greater brainstem representation of pitch-related spectral amplitudes in a repetitive vs. a variable stimulus context was associated with better SIN perception. The ability to modulate representation of pitch based on stimulus regularities is important for “tagging” relevant speech features, a key component of object identification and SIN perception. Therefore, in addition to difficulties with pitch perception and subsequent auditory grouping, children with diminished F_0 and H_2 representation are likely to have reduced ability to lock onto the pitch of the target signal, contributing to poorer SIN performance.

Subcortical spectral magnitudes are likely reflected by an interplay of central and peripheral processes. Brainstem responses to speech are determined by the acoustics of the incoming signal, and peripheral sensory processes may be impacted by environmental noise and/or hearing loss. However, cognitive processes such as attention, memory and object formation likely also affect subcortical encoding of sound (Lukas, 1981; Bauer and Bayles, 1990; Galbraith et al., 1997, 1998; Parbery-Clark et al., 2009a). When a listener judges a signal as important, auditory attention works to extract relevant signal elements from the competing background noise and stores them in working memory (Johnson and Zatorre, 2005). The cortex then uses this information to make predictions regarding the most relevant features of the stimulus, likely resulting in corticofugal enhancement in lower structures (Gao and Suga, 2000; Ahissar and Hochstein, 2004; Musacchia et al., 2007; Wong et al., 2007; de Boer and Thornton, 2008; Nahum et al., 2008; Song et al., 2008; Chandrasekaran et al., 2009b; Hill and Miller, 2010). Subcortical enhancement of relevant features in turn provides improved signal quality to the auditory cortex. Thus children with poor SIN perception likely have deficient encoding of sound due in part to a failure of cognitively-based processes such as attention, memory, and noise exclusion to shape effective subcortical sensory function. The differences in F_0 representation were seen only in response to the time-varying transition region of the stimulus, likely reflecting a selective enhancement of this response

component rather than an overall gain effect. This kind of top-down selective enhancement has also been proposed to explain differences between musicians and non-musicians’ subcortical responses to specific salient and behaviorally-relevant aspects of the signal (Lee et al., 2009; Strait et al., 2009; Chandrasekaran and Kraus, 2010a).

It would be useful to determine which children have biological signatures that reflect an underlying excessive difficulty hearing in background noise. Perhaps these children could benefit more from auditory training programs if the underlying issue (i.e., noise exclusion) was more directly addressed. Moreover, training gains should transfer to typical learning environments that tend to be noisy (e.g. classrooms). Consequently, some children may benefit from speech-in-noise training, training to strengthen the detection of relevant signals from complex soundscapes (e.g. music) and/or the use of augmentative communication devices (e.g., wireless assistive listening devices that provide an enhanced SNR). Objective neural indices, such as auditory brainstem responses, can aid the development and ongoing assessment of individual-specific training approaches that are needed to address the heterogeneous disorders associated with difficulty hearing speech in noise.

Acknowledgements

This work was funded by the National Institutes of Health (RO1 DC01510) and the Hugh Knowles Center of Northwestern University. We would like to thank Trent Nicol, Judy Song, Jane Hornickel, Alexandra Parbery-Clark, Jennifer Krizman, and Kyung Myun Lee for their helpful comments and suggestions regarding an earlier version of the manuscript. We would especially like to thank the children and their families who participated in the study.

References

- Abdo, A.G.R., Murphy, C.F.B., Schochat, E., 2010. Habilidades auditivas em crianças com dislexia e transtorno do déficit de atenção e hiperatividade. *Pró-Fono Revista de Atualização Científica* 22, 25–30.
- Ahissar, M., 2007. Dyslexia and the anchoring-deficit hypothesis. *Trends Cogn. Sci.* 11, 458–465.
- Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464.

- Ahissar, M., Lubin, Y., Putter-Katz, H., Banai, K., 2006. Dyslexia and the failure to form a perceptual anchor. *Nat. Neurosci.* 9, 1558–1564.
- Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47.
- Alain, C., Arnott, S.R., Picton, T.W., 2001. Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psych.-Human Perception Perform.* 27, 1072–1089.
- Alain, C., Reinke, K., He, Y., Wang, C., Lobaugh, N., 2005. Hearing two things at once: neurophysiological indices of speech segregation and identification. *J. Cogn. Neurosci.* 17, 811–818.
- Anderson, S., Kraus, N. in press. Objective neural indices of speech-in-noise perception. *Trends in Amplif.*
- Anderson, S., Skoe, E., Chandrasekaran, B., Kraus, N., 2010. Neural timing is linked to speech perception in noise. *J. Neurosci.* 30, 4922–4926.
- ANSI, 2002. Acoustical performance criteria, design requirements and guidelines for schools. ANSI S12.60.
- Assmann, P.F., Summerfield, Q., 1987. Perceptual segregation of concurrent vowels. *J. Acoust. Soc. Am.* 82, S120.
- Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., Kraus, N., 2009. Reading and subcortical auditory function. *Cereb. Cort.* 19, 2699–2707.
- Banai, K., Nicol, T., Zecker, S.G., Kraus, N., 2005. Brainstem timing: implications for cortical processing and literacy. *J. Neurosci.* 25, 9850–9857.
- Basu, M., Krishnan, A., Weber-Fox, C., 2010. Brainstem correlates of temporal auditory processing in children with specific language impairment. *Dev. Sci.* 13, 77–91.
- Bauer, L.O., Bayles, R.L., 1990. Precortical filtering and selective attention: an evoked potential analysis. *Biol. Psychol.* 30, 21–33.
- Bench, J., Kowal, A., Bamford, J., 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br. J. Audiol.* 13, 108–112.
- Bird, J., Darwin, C.J., 1998. Effects of a difference in fundamental frequency in separating two sentences. In: Palmer, A.R., Rees, A., Summerfield, A.Q., Meddis, R. (Eds.), *Psychophysical and Physiological Advances in Hearing*. Whurr, London.
- Bradlow, A.R., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. *J. Speech Lang. Hear. Res.* 46, 80–97.
- Bregman, A.S., 1990. *Auditory Scene Analysis*. Cambridge, MIT.
- Brox, J.P., Nootboom, S., 1982. Intonation and the perceptual separation of simultaneous voices. *J. Phonetics* 10, 23–26.
- Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T.G., Kraus, N., 2009a. 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., 2010a. Music, noise-exclusion, and learning disabilities. *Music Perception* 47, 297–306.
- Chandrasekaran, B., Kraus, N., 2010b. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246.
- Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2009b. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang.* 108, 1–9.
- Culling, J.F., Darwin, C.J., 1993. Perceptual separation of simultaneous vowels: within and across-formant grouping by f_0 . *J. Acoust. Soc. Am.* 93, 3454–3467.
- Cunningham, J., Nicol, T.G., Zecker, S.G., Bradlow, A.R., Kraus, N., 2001. Neurobiological responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin. Neurophysiol.* 112, 758–767.
- de Boer, J., Thornton, 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. *J. Neurosci.* 28, 4929–4937.
- de Cheveigne, A., 1997. Concurrent vowel identification. III. A neural model of harmonic interference cancellation. *J. Acoust. Soc. Am.* 101, 2857–2865.
- Fellowes, J., Remez, R., Rubin, P., 1997. Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics* 59, 839–840.
- 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., Bhuta, S.M., Choate, A.K., Kitahara, J.M., Mullen, T.A.J., 1998. Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9, 1889–1893.
- Galbraith, G.C., Jhaveri, S.P., Kuo, J., 1997. Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition. *Electroencephalogr. Clin. Neurophysiol.* 102, 46–53.
- Gao, E., Suga, N., 2000. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: Role of the corticofugal system. *PNAS* 97, 8081–8086.
- Gorga, M., Abbas, P., Worthington, D., 1985. Stimulus calibration in ABR measurements. In: Jacobsen, J. (Ed.), *The Auditory Brainstem Response*. College Hill Press, San Diego.
- Hall, J.W., Mueller, H.G., 1997. *Audiologists' Desk Reference*. Singular, San Diego.
- Heinrich, A., Schneider, B.A., Craik, F.I., 2007. Investigating the influence of continuous babble on auditory short-term memory performance. *Q. J. Exp. Psychol.* 61, 735–751.
- Hill, K.T., Miller, L.M., 2010. Auditory attentional control and selection during cocktail party listening. *Cereb. Cortex* 20, 583–590.
- Hornickel, J., Skoe, E., Nicol, T., Zecker, S., Kraus, N., 2009. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc. Nat. Acad. Sci.* 106, 13022–13027.
- Johnson, J.A., Zatorre, R.J., 2005. Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cereb. Cortex* 15, 1609–1620.
- Johnson, K.L., Nicol, T., Zecker, S.G., Bradlow, A.R., Skoe, E., Kraus, N., 2008. Brainstem encoding of voiced consonant-vowel stop syllables. *Clin. Neurophysiol.* 119, 2623–2635.
- Klatt, D., 1980. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* 67, 971–995.
- Knecht, H.A., Nelson, P.B., Whitelaw, G.M., Feth, L.L., 2002. Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements. *Am. J. Audiol.* 11, 65.
- Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Kraus, N., Nicol, T.G., 2005. How can the neural encoding and perception of speech be improved? In: Merzenich, M., Syka, J. (Eds.), *Plasticity and Signal Representation in the Auditory System* (New York: Springer).
- 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. *Ann. N. Y. Acad. Sci.* 1169, 543–557.
- Krishnan, A., Gandour, J.T., 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain Lang.* 110, 135–148.
- Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cog. Brain Res.* 25, 161–168.
- Lagace, J., Jutras, B., Gagne, J., 2010. Auditory processing disorder and speech perception problems in noise: finding the underlying origin. *Am. J. Audiol.* 1, 1059.
- Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* 29, 5832–5840.
- Lin, J.-Y., Hartmann, W.M., 1998. The pitch of a mistuned harmonic: evidence for a template model. *J. Acoust. Soc. Am.* 103, 2608–2617.
- Lukas, J.H., 1981. The role of efferent inhibition in human auditory attention: an examination of the auditory brainstem potentials. *Int. J. Neurosci.* 12, 137–145.
- Meddis, R., O'Mard, L., 1997. A unitary model of pitch perception. *J. Acoust. Soc. Am.* 102, 1811–1820.
- Moore, B.C., Peters, R.W., Glasberg, B.R., 1985. Thresholds for the detection of inharmonicity in complex tones. *J. Acoust. Soc. Am.* 77, 1861–1867.
- Moore, D.R., 2007. Auditory processing disorders: acquisition and treatment. *J. Commun. Dis.* 40, 295–304.
- Moore, D.R., Rosenberg, J.F., Coleman, J.S., 2005. Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. *Brain Lang.* 94, 72–85.
- Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr. Clin. Neurophysiol.* 35, 665–667.
- Musacchia, G., Arum, L., Nicol, T., Garstecki, D., Kraus, N., 2009. Audiovisual deficits in older adults with hearing loss: biological evidence. *Ear Hear.* 30, 505–514.
- Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Nat. Acad. Sci.* 104, 15894–15898.
- Nahum, M., Nelken, I., Ahissar, M., 2008. Low-level information and high-level perception: the case of speech in noise. *PLoS Biol.* 6, 978–991.
- Oxenham, A.J., 2008. Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends Amplif.* 12, 316–331.
- Parbery-Clark, A., Skoe, E., Kraus, N., 2009a. 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., 2009b. Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.
- Pichora-Fuller, M.K., Souza, P.E., 2003. Effects of aging on auditory processing of speech. *Int. J. Audiol.* 42, S11–S16.
- Russo, N., Nicol, T., Musacchia, G., Kraus, N., 2004. Brainstem responses to speech syllables. *Clin. Neurophysiol.* 115, 2021–2030.
- Scheffers, M.T.M., 1983. Simulation of auditory analysis of pitch: an elaboration on the DWS pitch meter. *J. Acoust. Soc. Am.* 74, 1716–1725.
- Shield, B.M., Dockrell, J.E., 2003. The effects of noise on children at school: a review. *Building Acoust.* 10, 97–116.
- Shield, B.M., Dockrell, J.E., 2008. The effects of environmental and classroom noise on the academic attainments of primary school children. *J. Acoust. Soc. Am.* 123, 133–144.
- Shinn-Cunningham, B.G., Best, V., 2008. Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299.
- Skoe, E., Kraus, N., 2010. Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324.
- Song, J.H., Banai, K., Russo, N.M., Kraus, N., 2006. On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol. Neuro-Otol.* 11, 233–241.
- Song, J.H., Nicol, T. & Kraus, N. Test-retest reliability of the speech-evoked auditory brainstem response in young adults. *Clin. Neurophysiol.* in press.
- Song, J.H., Skoe, E., Banai, K. & Kraus, N. Perception of speech in noise: neural correlates. *J. Cogn. Neurosci.* in press, doi:10.1162/jocn.2010.21556.
- Song, J.H., Skoe, E., Wong, P.C.M., Kraus, N., 2008. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902.
- Sperling, A.J., Zhong-Lin, L., Manis, F.R., Seidenberg, M.S., 2005. Deficits in perceptual noise exclusion in developmental dyslexia. *Nat. Neurosci.* 8, 862–863.
- 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. *Eur. J. Neurosci.* 29, 661–668.

- Summers, V., Leek, M.R., 1998. F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J. Speech Lang. Hear. Res.* 41, 1294–1306.
- Tallal, P., Piercy, M., 1974. Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia* 12, 83–93.
- Tallal, P., Stark, R.E., 1981. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. *J. Acoust. Soc. Am.* 69, 568–574.
- Tzounopoulos, T., Kraus, N., 2009. Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62, 463–469.
- Wible, B., Nicol, T., Kraus, N., 2004. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol. Psychol.* 67, 299–317.
- Wible, B., Nicol, T., Kraus, N., 2005. Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain* 128, 417–423.
- Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Woodcock, R.W., McGrew, K.S., Mather, N., 2001. *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca. Riverside Publishing, IL.
- Zhu, J., Garcia, E., 1999. *The Weschler Abbreviated Scale of Intelligence (WASI)*. Psychological Corporation, New York.
- Ziegler, J.C., Pech-Georgel, C., George, F., Alario, F.-X., Lorenzi, C., 2005. Deficits in speech perception predict language learning impairment. *Proc. Nat. Acad. Sci.* 102, 14110–14115.
- Ziegler, J.C., Pech-Georgel, C., George, F., Lorenzi, C., 2009. Speech-perception-in-noise deficits in dyslexia. *Dev. Sci.* 12, 732–745.