Brain Stem Response to Speech: A Biological Marker of Auditory Processing

Krista L. Johnson, Trent G. Nicol, and Nina Kraus

The auditory brain stem response to speech mimics the acoustic characteristics of the speech signal with remarkable fidelity. This makes it possible to derive from it considerable theoretical and clinically applicable information relevant to auditory processing of complex stimuli. Years of research have led to the current characterization of these neural events with respect to the underlying acoustic information they reflect. The majority of data reviewed here originates from studies using a /da/ syllable to elicit the brain stem response, which consists of transient and periodic (frequency following) neural activity. We describe how the human auditory brain stem response separately encodes source and filter characteristics of the acoustic signal, which reflect paralinguistic and linguistic information simultaneously conveyed in speech. In normal-hearing individuals, these two classes of response components (source and filter) are highly correlated within a class but not between classes. This response dissociation becomes pronounced when stimuli are presented in background noise or with faster stimulus rates. In addition, some learning-impaired children show a selective deficiency in the neural encoding of acoustic features associated with the filter characteristics of speech. These children show no deficits in the encoding of source components, further supporting the notion of separate neural mechanisms. Overall, the auditory brain stem response to speech provides a way to access subcortical auditory processing mechanisms and may be used as a biological marker of deficient sound encoding associated with learning and auditory processing disorders.

(Ear & Hearing 2005;26;424–434)

Sensory experiences of all types contribute to how an organism will react to the surrounding environment. For example, a distinct odor will warn a predator not to attack a skunk; a moving shadow will allow a hawk to spot prey; touching a hot pan will warn a child of danger. Appropriate reaction is an important outcome to almost all events, and sensory systems are especially equipped to respond well to rapidly occurring stimuli exhibiting distinct temporal features. Exposure to sensory stimuli from all modalities is important from an evolutionary standpoint. However, there is one exceptional stimulus which, although not mandatory for survival, is an essential part of everyday life: speech. The scope of this review is to consider how the subcortical auditory system preconsciously encodes the building blocks of what will become the acoustic message (word) conveyed by spoken language and how this encoding provides a biological marker of auditory processing.

Speech is a complex signal varying in many acoustic dimensions continuously over time. Unlike printed words, conversational speech has more subtle markers that cue a listener to the beginning and end of meaningful segments. How do listeners encode such rapid, brief, and complex stimuli into meaningful units?

Brain Stem Response to Speech Sounds

Recording brain stem response to sound has long been established as a valid and reliable means to assess the integrity of the neural transmission of acoustic stimuli. Transient acoustic events induce a pattern of voltage fluctuations in the brain stem resulting in a familiar waveform, yielding information about brain stem nuclei along the ascending central auditory pathway (for review, see Hood, 1998; Jacobson, 1985). An accurate manifestation of stimulus timing in the auditory brain stem is a hallmark of normal perception (Sininger & Starr, 2001). Disruptions in this systematic progression on the order of fractions of milliseconds are clinically significant in the diagnosis of hearing loss and brain stem pathology. With a sensory system so attuned to the temporal characteristics of sound, it is logical to ask how brain stem timing contributes to neural encoding of speech.

Animal models have been used to describe auditory nerve and cochlear nucleus single-unit response properties for synthetic speech-like sounds (Delgutte, 1984; Delgutte & Kiang, 1984a, 1984b, 1984c, 1984d; Young & Sachs, 1979). Not only do auditory nerve and cochlear nucleus fibers show increased phase-locked activity to the formant harmonics in the stimulus, but separate populations of neurons

Auditory Neuroscience Laboratory (K.L.J., T.G.N., N.K.), Departments of Communication Sciences (K.L.J., T.G.N., N.K.), Institute for Neuroscience (N.K.), Neurobiology and Physiology (N.K.), and Otolaryngology (N.K.), Northwestern University, Evanston, Illinois.

appear to encode the first and second formant. Neural encoding of speech in more rostral structures such as the lateral lemniscus and inferior colliculus has not been studied extensively. Moreover, based on the phase-locking limitations of these structures, it is assumed that neural encoding of the periodic acoustic properties of speech at such rostral areas would be limited to temporal events well below the second formant.

Speech stimuli have also been used in humans to study the response characteristics of the frequency following response (FFR) (Galbraith et al., 2004; Galbraith, Arbagey, Branski, Comerci, & Rector, 1995; Krishnan, 2002; Krishnan, Xu, Gandour, & Cariani, 2004). The FFR arises from the harmonic portion of the stimulus and is characterized as a series of transient neural events phase-locked to periodic information within the stimulus (Batra, Kuwada, & Maher, 1986; Marsh & Worden, 1968; Sohmer & Pratt, 1977). Galbraith et al. (1995) demonstrated that the FFR elicited by word stimuli reflects the stimulus accurately enough to allow it to be recognized as intelligible speech when "played back" as an auditory stimulus. More recently, Galbraith and colleagues (2004) have suggested that based on the FFR pattern of activation for forward and backward speech, synaptic processing at the level of the brain stem is more effective for forward speech stimuli, characterized by highly familiar prosodic and phonemic structure, than to backward speech. Krishnan has studied the FFR elicited by synthetic vowels to relate phase-locking characteristics of brain stem neurons to individual harmonics of a complex sound. Results suggest that human FFR spectra show clear and distinct peaks corresponding to formant frequencies of steady-state synthetic vowels (Krishnan, 2002). Subsequent studies by Krishnan and colleagues used Chinese syllables to show that pitch representation in the auditory brain stem is based on temporal patterns of phase-locked neural activity of the fundamental frequency, as represented by the FFR (Krishnan et al., 2004).

Our laboratory has dedicated significant effort to understanding how the brain stem response neurally represents speech syllables (Banai, Nicol, Zecker, & Kraus, 2005; Cunningham, Nicol, Zecker, & Kraus, 2001; Hayes, Warrier, Nicol, Zecker, & Kraus, 2003; King, Warrier, Hayes, & Kraus, 2002; Kraus & Nicol, 2005; Russo, Nicol, Musacchia, & Kraus, 2004; Russo, Nicol, Zecker, Hayes, & Kraus, 2005; Wible, Nicol, & Kraus, 2004a, 2004b, 2005). A primary focus has been to understand the neural correlates of specific acoustic events within a speech syllable. This work has also led to the translation of

findings into the clinic with the development of BioMAP (Biological Marker of Auditory Processing, distributed by Bio-logic Systems Corporation) that will allow clinicians to identify disordered auditory processing of sound that has been associated with learning impairments in many children.

Stimulus and Recording Parameters

With the exception of Cunningham et al. (2001), who used a slightly modified /da/ stimulus, the following parameters are common to the work from our laboratory reviewed here. A Klatt cascade/parallel formant synthesizer (Klatt, 1980) was used to synthesize a 40-msec speech-like /da/ syllable at a sampling rate of 10 kHz (Fig. 1). The stimulus was constructed to include an onset burst frication at F_3, F_4, and F_5 during the first 10 msec, followed by 30-msec F_1 and F_2 transitions ceasing immediately before the steady-state portion of the vowel. Although the stimulus does not contain a steady-state portion, it is psychophysically perceived as a consonant-vowel speech syllable. We chose a stimulus that was short enough to minimize test time while still containing key acoustic phonetic information. Test stimuli were delivered to the right ear through Etymotic ER-3 earphones at an intensity of 80 dB SPL whereas the left ear was unoccluded.

Brain stem evoked potentials were obtained to randomly presented alternating polarity stimuli. To extract the neural response from the cochlear microphonic and eliminate stimulus artifact, the alternating polarities were added together (Gorga, 1985). Responses were differentially recorded from Cz-to-ipsilateral earlobe, with the forehead as ground. This montage accentuates more rostral (i.e., lateral lemniscus/inferior colliculus) components rather than more peripheral (i.e., auditory nerve/cochlear nucleus) contributions (Galbraith, 1994; Galbraith et al., 2000; Møller, Jannetta, & Sekhar, 1988). Three blocks of 1000 responses to each polarity of /da/ were collected with a 51-msec interstimulus interval. A 60-msec recording window (including a 10-msec prestimulus period) was used. Responses were sampled at 20,000 Hz and bandpass filtered on-line from 100 to 2000 Hz, using a 6 dB/octave filter roll-off. Although the poststimulus response epoch is long enough to contain cortical contributions, the response reflects largely events of brain stem origin due to the frequency content of the peaks, which are higher than would be seen in a cortical or middle latency generated response (Kavanagh, Harker, & Tyler, 1984; Sayers, Beagley, & Henshall, 1974; Suzuki, Kobayashi, & Hirabayashi, 1983; Yamamoto, Sakabe, & Kaiho, 1979).
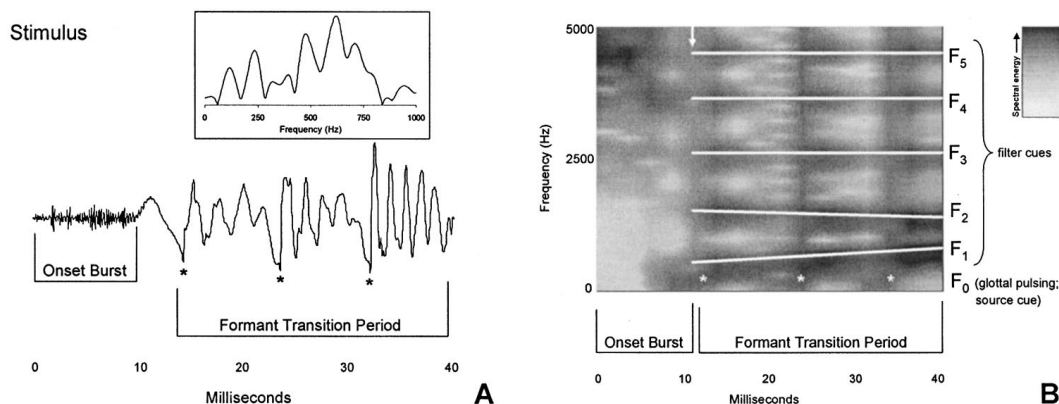
Fig. 1. (A) Bottom: Time-amplitude waveform of the synthesized speech stimulus /da/. The first 10 msec contains the onset burst of the consonant /d/. The following 30 msec is the formant transition to the vowel /a/. Asterisks mark the period of the fundamental frequency (~120 Hz). Top: Frequency spectrum over the periodic portion of the stimulus (17 to 38 msec). (B) Spectrogram of /da/. Darker areas represent regions of highest spectral energy. The 10-msec onset burst contains diffuse high-frequency energy. Arrow represents release of the glottal air pressure and the initial frequency information of F_1, F_2, and F_3. The 30-msec formant transition period represents the formant frequency transitions (marked with white lines) to the intended vowel. Initial frequencies of F_2 and F_3 provide information relevant to consonant identity. Relative spacing between the frequencies of F_1 and F_2 at the conclusion of the transition provides information relevant to vowel identity. Asterisks mark glottal pulsing, visible as vertical bands in the spectrum, representing the F_0 of the stimulus.

The neural response to the speech syllable /da/ can be described morphologically in terms of an onset response and an FFR, as seen in Figure 2. The robust onset response is similar to that observed in response to a tone or click stimulus, consisting of waves I, III, and the VA complex. The voiced portion of the stimulus evokes the periodic portion of the response, the FFR, which reflects phase-locking to the waveform of the stimulus.

An initial step in understanding the functional relationship between the acoustic structure of speech and the brain stem response to speech was to establish a valid and reliable means by which to characterize the overall neural activity of the brain stem in response to the speech sound /da/ (Cunningham et al., 2001; Russo et al., 2004). Because rapid temporal changes and complex spectral distributions are inherent in speech, both micromeasures (transient) and macromeasures (sustained) of timing and magnitude are used to describe the response. Timing measures provide insight into (1) the accuracy with which brain stem nuclei synchronously respond to acoustic stimulation (e.g., peak latency, interpeak interval, and slope) and (2) the fidelity with which the response mimics the stimulus or the degree to which it is degraded by background noise (e.g., stimulus-to-response correlations and quiet-to-noise response correlations). Magnitude measures provide information about (1) the robustness with which the brain stem nuclei respond to acoustic stimulation (e.g., peak amplitude and root-mean-square amplitude of activation) and (2) the size of spectral components within the response (e.g., frequency-domain analysis). Response replicability and test-retest reliability were established in both quiet and background noise conditions (Russo et al., 2004). Although the speech-evoked response can be analyzed in all of the ways listed above, Table 1 illustrates those measures that have proven to be the most informative, reliable, and clinically relevant. Note that transient response measures of peak latencies and VA slope are highly detectable with little variability.
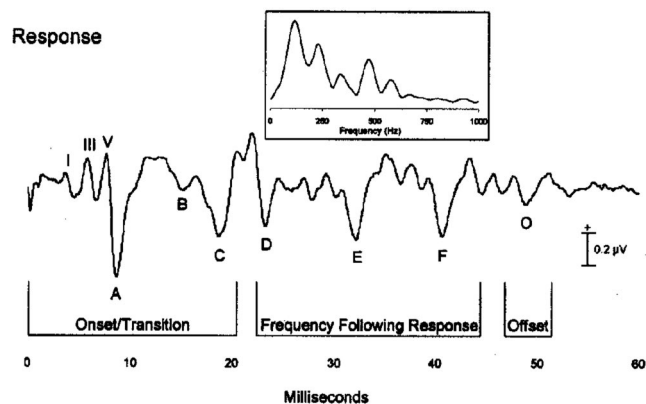

Fig. 2. Auditory brain stem response to the speech syllable /da/ (mean of 24 normal-hearing subjects). Waves I, III, and V are standard nomenclature for the onset response complex. The remaining transient peaks A, B, C, D, E, F, and O comprise the brain stem response to speech (peak B is inconsistently present and therefore is not considered further). Top: Frequency spectrum over the periodic portion of the response (23 to 44 msec).

TABLE 1. Clinically relevant auditory brain stem measures (n = 88)

Transient response measures			Sustained response measurements	
Peak Latency	Detection (%)	SD	Timing	Magnitude
V	100	0.25	Stimulus-to-response correlation	F0 amplitude
A	100	0.31	Quiet-to-noise response correlation	F1 amplitude
C	100	0.47		
D	95	0.43		
E	98	0.44		
F	100	0.42		
O	100	0.52		
VA slope	100	0.34		

Brain Stem Encoding of Speech Characteristics: A Conceptual Framework

Acoustic Characteristics of the Syllable /da/

The acoustic characteristics of speech supply a listener with cues enabling identification of both the phonetic content of the message as well as information pertaining to who is speaking and the intention of the message. Linguistic information is necessary to distinguish the meaning of the message (consonants and vowels); paralinguistic information conveys the intention, or how the message is expressed (e.g., statement versus question, angry versus happy emotional state). Paralinguistic acoustic elements add a multidimensional aspect to speech that is separate from the phonetic information of the verbal message. Acoustically, these cues are conveyed by "source" and "filter" characteristics. The source-filter model of speech production states that speech comprises (i) vibration of the vocal folds reacting to airflow from the lungs (source) and (ii) the shape of the vocal tract and articulator manipulation of the oral cavity, tongue, lips, and jaw (filter) (Fant, 1960). Generally but not exclusively, paralinguistic information is conveyed by the source, whereas linguistic information is conveyed by particular filter shapes.

Acoustically, the source of the utterance /da/ is the fundamental frequency (F_0), in this case ramping from 100 Hz to 120 Hz. The filter is manifested acoustically by (i) the initial frication (tongue at the roof of the mouth blocking airflow, followed by a release of the blockage) resulting in the 'd' sound, and (ii) a shift in the articulators (lowering of the tongue and widening of the mouth) transforming the initial "d" into "ah". This filtering of the source selectively accentuates certain harmonics of the fundamental, resulting acoustically in prominent peaks in the spectrum, the formants.

Parallels in Stimulus and Response Morphology

A visual analysis of the /da/ stimulus waveform and its corresponding brain stem response reveals several similarities. Shifting the stimulus waveform by approximately 7 msec to account for neural conduction time reveals an even more striking match (Fig. 3A). **F3**

The most evident parallel is the comparable periodic nature of the glottal pulsing of the stimulus with the three major peaks within the FFR. The period between response peaks D, E, and F corresponds to the wavelength of the F_0 of the utterance. These peaks represent the glottal pulsing of the vocals folds and are thus representing source information. Moreover, Fourier analysis of this portion of the response confirms a spectral peak at the frequency of F_0. Figure 3B illustrates the spectral similarity between the response over the isolated time range encompassing peaks D, E, and F of the response and the voiced portion of the stimulus.

The remaining peaks within the auditory brain stem response can be thought to represent filter characteristics. Waves V, A, C, and O are events that occur in response to transient stimulus events separate from the periodic acoustic events in the stimulus. The VA complex reflects a highly synchronized neural response to the onset of the stimulus. Peak C probably is a response to the onset of the voicing that occurs at 10 msec after stimulus onset. Wave O probably is a response to the cessation of sound, as it corresponds temporally to the offset of the stimulus. Together, these transient peaks, the timing of which is sensitive to stimulus spectrum, comprise responses to the acoustic filter characteristics of the syllable. Additionally, the spacing of the small, higher-frequency fluctuations between waves D, E, and F correspond in frequency to the F_1 of the stimulus. F_2 is also an important acoustic cue for identifying linguistic content. However, F_2 and the higher formants in the /da/ are beyond the phase-
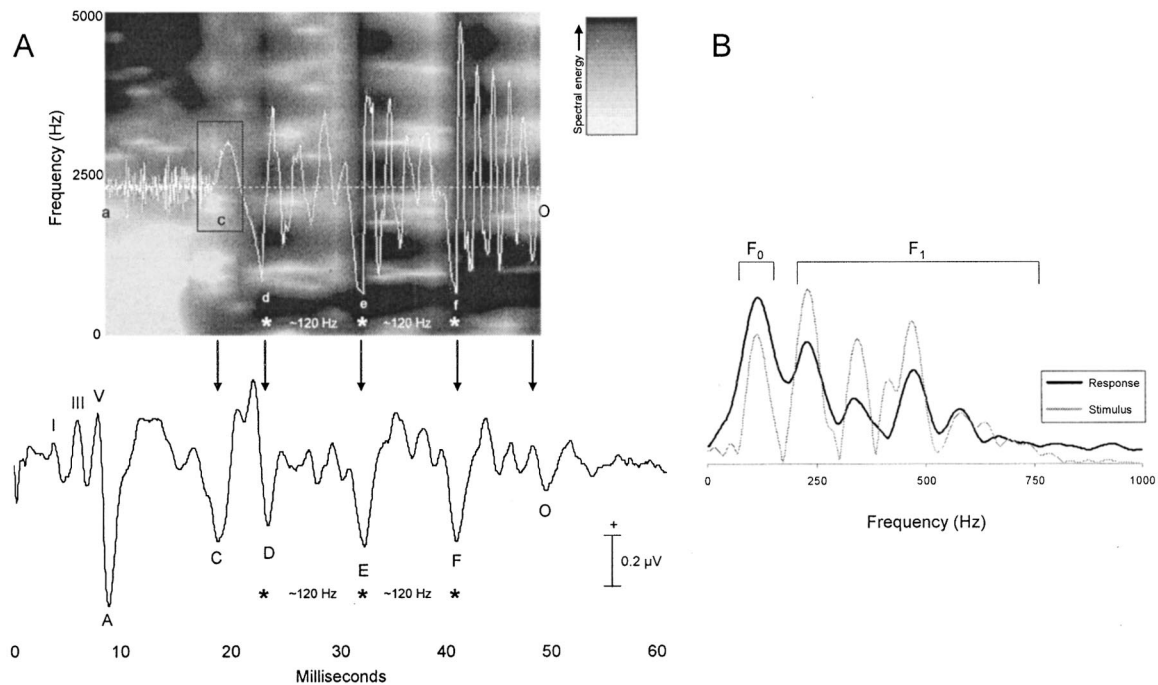
Fig. 3. A, Framework for how speech sound structure is represented by brain stem neurons. Neural events (uppercase letters) reflect a direct mapping of stimulus characteristics (lowercase letters). The stimulus waveform has been shifted 7 msec to compensate for neural lag in the response. The spectrogram and overlaid waveform illustrate the elements of the stimulus and corresponding peaks in the response. Note that the wavelengths between peaks d, e, and f (the F_0 of the stimulus) correspond to peaks D, E, and F of the response (marked by asterisks). Also note that waves C and O correspond to major stimulus feature changes (wave C: transition between onset burst and more periodic portion; wave O: stimulus offset). B, Fourier analysis of the stimulus (light line; filtered at 400 Hz to mimic the low-pass characteristics of the midbrain) and the brain stem response (dark line, time range of 23 to 44 msec). The brain stem response demonstrates remarkable fidelity to the stimulus in the frequency domain for both F_0 and F_1.

locking limit of the rostral brain stem (Blackburn & Sachs, 1989; Frisina, 2001; Frisina, Smith, & Chamberlain, 1990; Joris, Schreiner, & Rees, 2004; Langner & Schreiner, 1988; Wang & Sachs, 1994) and, consequently, are not evident in the response. Nonetheless, the frequency content of F_2-F_5 probably is represented by transient response latencies and amplitudes, thereby imparting additional linguistic information about the utterance to the response. Future studies will need to examine each one of these observations systematically.

Although source and filter characteristics occur simultaneously in the speech signal and in the response, specific components of the brain stem response reflect these acoustic characteristics separately. Thought of in this manner, the auditory brain stem response reflects encoding of both the linguistic and paralinguistic elements in speech. One neural "stream" represents the source characteristics of speech by encoding the sustained information of F_0, whereas another neural "stream" represents the filter characteristics by encoding the transient and/or rapid information of the formants, onsets, offsets, and other transitional events.

This framework coincides with how the auditory system encodes spectral and periodic information. Spectral coding probably is the primary analytic mode for filter cue extraction. The auditory brain stem is tonotopically organized (Langner, 1997; Langner, Sams, Heil, & Schulze, 1997), and it has been demonstrated that the cortical coding of stop consonants depends on the frequency content of the component formants (Martin, Sigal, Kurtzberg, & Stapells, 1997; McGee, Kraus, King, Nicol, & Carrell, 1996; Steinschneider, Schroeder, Arezzo, & Vaughan, 1993). On the other hand, it is periodicity, not frequency, which affords the quality of "pitch" to a sound, a phenomenon readily illustrated by "missing" fundamental stimuli. Thus, temporal encoding of F_0 probably is the mediating influence in the percept of source cues in the context of the present discussion. Systematic encoding of periodicity has also been demonstrated, and there is some evidence for the existence of orthogonal periodic and tonotopic maps in the cochlear nucleus, inferior colliculus, and auditory cortex (Langner, 1997; Langner et al., 1997; Langner & Schreiner, 1988; Merzenich, Knight, & Roth, 1975; Schreiner & Langner, 1988; Suga & O'Neill, 1979).

The data that drive the notion of a functional separation in the way that source and filter cues are represented in the brain stem response are reviewed below. They are derived from studies that have examined (I) correlations among response components, (II) the effects of noise and rate, (III) how responses are affected by learning impairment and (IV) auditory training, and (V) consideration of how brain stem encoding may be a precursor to the notion of cortical "what" and "where" pathways.

Functional Separation I: Response Correlations

In normal-hearing subjects, a correlational analysis of the various response components (Russo et al., 2004) supports a relationship between latencies of the VA complex and wave C, both of which are also related to the spectral magnitude of F_1 in the response. Similarly, the latencies of waves D, E, and F are highly correlated with one another and with the spectral magnitude of F_0 in the response. However, these two distinct groups of response characteristics are unrelated to one another. Consequently, it appears that elements of the neural mechanism encoding primarily source information about speech are different from the neural mechanism responsible for encoding primarily filter information.

Functional Separation II: Effects of Background Noise

Competing acoustic signals are a part of everyday life; a deeper understanding of low-level neural functioning of the brain stem in noise may contribute to understanding high-level/behavioral responses to speech in noise. Consonants in general and stop consonants such as /d/ in particular are highly susceptible to the ill effects of a noisy environment. Acoustically, stop consonants are characterized by low amplitude, stochastic, temporally rapid, and discrete events. Vowels, though linguistic, are "carried" on the source-based fundamental and are of relatively large amplitude. Because of this, vowels are naturally more resilient in noise.

Knowing how background noise degrades the perception of speech, it follows that a particular pattern of degradation in the neural response would be predicted. Russo et al. (2004) demonstrated that noise has different effects on the brain stem encoding of /da/, such that the components associated with the filter were degraded more severely than those signaling the source. Most notably, the onset responses V and A were highly compromised in background noise. Equally interesting is that peak F of the FFR remained temporally stable and easily identifiable in noise. Last, the F_0 spectral representation in the response was resilient to the effects of noise, whereas the F_1 information represented in the response was attenuated considerably.

Functional Separation III: Learning Impairment

A growing body of literature that has evaluated speech-evoked brain stem response differences between normal-learning children (NL) and children with learning problems (LP) suggests that brain stem measures relating to the encoding of filter information can serve as a biological marker for brain stem neural asynchrony in children with language-based learning problems, such as dyslexia, or children diagnosed with central auditory processing disorders. A consistent finding is that about one third of LP children exhibit a unique pattern of auditory neural activity that distinguishes them from the larger LP population (Banai et al., 2005; Cunningham et al., 2001; Hayes et al., 2003; King et al., 2002; Russo et al., 2005; Wible et al., 2004a, 2004b, 2005). These children exhibit delayed peak latency or shallower slope measures of the VA onset complex and of waves C and O, indicating poor synchrony to transient events (Fig. 4A). It was also **F4** found that LP children displayed diminished F_1 spectral content, another filter cue (Fig. 4B).

Environmental stresses such as noise and rapidly presented stimuli can further negatively influence neural encoding of linguistic information in LP children. In the presence of background noise or rapid stimulation, spectral cues present in the F_1 but not F_0 region were diminished in the LP population (Cunningham et al., 2001; Wible et al., 2004a). Recently, in guinea pigs, trains of rapidly presented speech syllables in noise were used to "stress" the normal system to simulate the particular response patterns demonstrated by LP children. It was found that neural activity in the inferior colliculus relating to F_1 but not F_0 demonstrated sensitivity to rapid stimulation, consistent with the impaired responses seen in LP children (Wible et al., 2004b).

There are also cortical consequences related to an abnormal brain stem response to /da/. Wible et al. (2004a) demonstrated that there was a relationship between VA slope of the brain stem and the cortical resistance to the degrading effects of background noise. He showed that the broader the slope, the more susceptible the cortical response (elicited to the same syllable) was to degradation in noise. Moreover, Banai and colleagues (2005) showed that LP children with abnormal speech-evoked brain stem onset responses also have significantly reduced mismatch negativity compared with NL children. **AQ: 1**
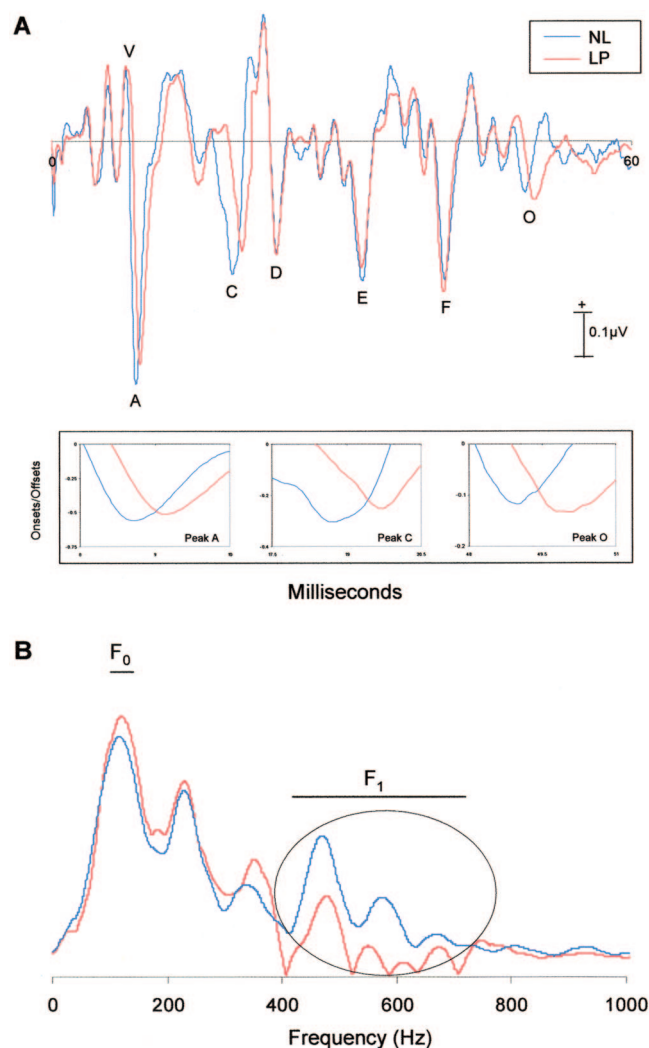
Fig. 4. Schematic summary of speech-evoked brain stem deficits that have been associated with LP children. Note: LP children do not necessarily show each manifestation of abnormal encoding illustrated here. (A) Schematic of onset/offset peak latency deficits typically observed in children with language-based learning problems as compared with NL children. LP children typically demonstrate a latency lag for one or all of peaks A, C, or O but not for peaks D, E, or F. B, Schematic of a response FFT over a time period of 23 to 44 msec in NL and LP children. LP children often demonstrate a decreased amplitude of F_1 (420 to 720 Hz) but not F_0 (103 to 121 Hz).

These studies suggest that abnormal speech encoding as low, as the brain stem may have broad neural encoding consequences throughout the entire auditory pathway.

LP children with deficits in brain stem measures relating to the encoding of filter information also perform below their counterparts on perceptually challenging auditory processing tasks. Performance on a speech sound perception task requiring discrimination between two stop consonants has con-

sistently been found to be poorer in LP children with delayed wave A and C latencies and diminished spectral information in the F_1 region (Cunningham et al., 2001; King et al., 2002; Wible et al., 2004a). A similar relationship was found between filter responses and an auditory backward masking task (Johnson, Nicol, Zecker, Wright, Kraus, 2004). Furthermore, children with brain stem timing deficits perform poorly on reading and spelling measures. Notably, brain stem timing deficits can predict literacy difficulties at the individual level (Banai et al., 2005).

LP children do not exhibit pan-auditory neural deficits; the NL and LP groups do not differ in their brain stem responses to source cues in speech. That is, it is not the case that brain stem responses are weaker across the board. The peaks within the sustained FFR that are locked to the periodicity of the stimulus F_0 do not differ between NL and LP children; waves D, E, and F do not show the latency delays in the LP population that are observed with peaks A, C, and O. Furthermore, spectral information within the F_0 region of the response remains robust (Fig. 4).

LP children appear to exhibit abnormal representation of specific neural activity rather than an overall deficit in neural synchrony. The pattern of deficits observed in the LP population is comparable, but to an enhanced degree, to what has been observed in the normal system (and in the guinea pig inferior colliculus) under stressful listening situations. It appears that the neural encoding challenges experienced by LP children are specific to processing the filter characteristics of sound and that the normal system experiences these challenges only under extreme conditions. A distinct pattern of inadequate representation of transient and temporally rapid activity and adequate representation of low frequency and sustained activity begins to emerge. Perhaps the neural timing deficits reported here lie on the milder side of a continuum of synchrony disorders, the most extreme being auditory neuropathy (Sininger & Starr, 2001).

Functional Separation IV: Response to Auditory Training

It is noteworthy that LP children with abnormal brain stem timing of peaks A and C are most likely to show both physiological and behavioral improvements after auditory training with commercially available software (Hayes et al., 2003; King et al., 2002; Russo et al., 2005). Thus, the brain stem response to speech can serve to inform recommendations of treatment strategies by providing an

objective indication that a child is likely to benefit from an auditory training program.

With respect to the plasticity of the brain stem response itself, training-related changes have been observed in the FFR but not the onset response (Russo et al., 2005). After training, FFRs to speech presented in background noise became more robust and better synchronized. The improvements seen with training can be viewed as reflecting more accurate neural encoding of filter information (neural activity relating to F_1) because the source information (neural activity relating to F_0) remained stable. It is therefore possible that auditory training has the effect of making aggregate brain stem neural activity less susceptible to the detrimental effects of background noise.

LP children who completed auditory training also showed improved cortical responses to speech syllables in noise (Hayes et al., 2003; King et al., 2002; Russo et al., 2005; Warrier, Johnson, Hayes, Nicol, & Kraus, 2004). These LP children also improved on a behavioral speech perception task (King et al., 2002) and tests of phoneme decoding and literacy (Hayes et al., 2003). Corticofugal descending pathways probably play a role in this subcortical plasticity after auditory training. Intercranial recordings in human auditory cortex have observed cortical activation as early as 12 msec in response to voiced speech sounds (Steinschneider, Volkov, Noh, Garell, & Howard, 1999). Thus, it can be theorized that cortical feedback may regulate the preconscious alterations of speech encoding in the auditory brain stem beyond this time frame. The effects of training on brain stem activity were primarily reflected in improved precision of the neural response recorded in background noise (Russo et al., 2005). The interpretation of auditory cortical modulation of brain stem activity is consistent with the aforementioned improvement observed in cortical responses in noise.

Functional Separation V: Cortical "What" and "Where" Pathways

Sensory processing streams, popularly called "what" and "where" pathways, are well established in the visual system as a neural scheme for separately carrying different facets of visual objects, namely, their identity and their position/motion to the cortex (Mishkin, Ungerleider, & Macko, 1983). A similar functional organization has been postulated in the auditory system (Belin & Zatorre, 2000; Kaas & Hackett, 1999; Rauschecker & Tian, 2000; Romanski et al., 1999). We have argued that the source and filter brain stem response classes considered here are early manifestations of auditory "what" and "where" streams in the cortex and represent speak-

er-as-auditory object and the building blocks of the message, respectively (Kraus & Nicol, 2005). Thus, it seems that before the what/where separation occurs in the cortex, there is a precursor of this separation earlier in the sensory pathway that is evident in the brain stem response to speech.

Further Research and Clinical Applications

A significant gap in our knowledge of how the brain stem responds to speech sounds still exists due to the limited repertoire of speech sounds used to date. Future research should include an expanded selection of speech stimuli, both synthetic and natural, as well various presentation levels and rates. Considerable advances in our understanding of the exact contribution brain stem neurons play in encoding speech will be gained from studies that incorporate a larger collection of syllables that vary systematically in both source and filter characteristics. Nevertheless, research thus far has provided a conceptual framework for considering the brain stem response with respect to the acoustic characteristics of the speech signal.

Brain stem responses to speech sounds probably have other applications, both in research and in clinical settings, when central auditory function is of interest. Applications include contributing information about the nature of a specific learning deficit in LP children, monitoring auditory training progress, delineating the effects of aging, assessing central auditory processing deficits in hearing aid and cochlear implant users, and assessing rehabilitation strategies relevant to both the design and objective fitting of sensory aids.

Diagnosis and Remediation of Learning Problems • Existing normative data and research with clinical populations indicate that the brain stem response to speech is clinically viable for the identification of brain stem timing deficits in children with auditory processing and learning problems. Consequently, speech-evoked brain stem responses provide a biological marker for identifying children with preconscious, disordered auditory processing of sound. Moreover, they can be used to identify those children who are likely to benefit from auditory training programs as well as objectively monitoring the effects of training. Additional large-scale studies are necessary to further delineate the sensitivity and specificity of the brain stem response to speech with respect to diagnosis and remediation outcomes. Future research on other clinical populations who may exhibit neural encoding deficits to paralinguistic acoustic cues (e.g., autism) may also prove to be fruitful. Because the brain stem matures early, application in infants and preschool children at risk

for learning problems can be envisioned. The upcoming release of the BioMAP will make such diagnostic applications possible in clinical settings.

Central Auditory Processing Deficits With Aging • Hearing loss associated with aging, presbycusis, is a common complaint. With this type of hearing loss, individuals often find it harder to hear in noisy places and find themselves withdrawing socially because of the frustration of not understanding what is said. Although presbycusis is caused by cochlear hair cell damage, deeper knowledge of the negative effects this has on the central auditory processing stream is needed. Research using speech-evoked auditory brain stem responses may improve our understanding of how the neural representation of speech is being altered by a peripheral hearing impairment and may delineate which aspects of speech are most susceptible to neural degradation associated with aging. Clinicians could use this information to counsel patients on compensatory strategies unique to their individual processing deficits.

Hearing Aid and Cochlear Implant Users: Selection and (Re)Habilitation • Sensory aids may be optimally designed to allow for specific and specialized acoustic manipulations. Industries can benefit by designing circuitry capable of transmitting altered acoustic signals emphasizing flexibility beyond amplitude and filtering adjustments enabling control over variations in the transmission of the temporal characteristic of the signal. An understanding of how various central auditory processing deficits are expressed in the brain stem in terms of objective speech encoding could aid in designing hearing aid and cochlear implant processing strategies. It is not unreasonable to imagine using speech-evoked brain stem responses as a tool to aid in selecting the most appropriate hearing device, based on which elements of neural speech encoding are deficient.

Speech-evoked auditory brain stem responses can be used to monitor central auditory function for any number of habilitation/rehabilitation strategies. Since auditory training experience can promote neural changes within the brain stem, this window into neural function could be applied to investigate training/experience-related plasticity during acclimatization to a listening device. By monitoring the brain stem response to speech as neural restructuring occurs, adjustments to the signal processing could be made to compensate and complement these neural changes. Such programming would enable a patient to receive the most optimal signal for maximum hearing aid/cochlear implant benefit.

CONCLUSIONS

Subtle and temporally complex acoustic cues present in speech exist as a means for a listener to accurately and meaningfully encode speech segments. These inherent acoustic characteristics carry linguistic and paralinguistic information that contribute uniquely to either the meaning or intent of the message.

The auditory brain stem precisely encodes rapid acoustic information, evident in both the onset and FFR portions of the response. Measures of temporal precision and spectral content throughout the response can characterize the brain stem response to a representative speech syllable. By understanding the morphological components of the response, we can better understand how these components relate directly to the inherent acoustic structure of speech. Moreover, we can begin to determine how source and filter aspects of speech are represented by specific neural events.

Several lines of evidence support the notion that neural encoding of source and filter cues are mediated by different mechanisms. The response characteristics pertaining to filter cues in speech vary together and are separate from those corresponding to source cues, which also vary together. In the presence of adverse listening conditions, such as background noise and rapid stimulation, the neural representation of source information remains relatively intact, whereas the neural activity representing filter information becomes highly degraded. Additionally, some children with language-based learning problems demonstrate selective disruptions in brain stem encoding of the filter characteristics of the speech signal. Furthermore, the outcome of auditory training programs suggests that physiological improvements of neural mechanisms encoding some filter cues can be achieved independent of any changes in the neural mechanisms that encode source cues.

The theoretical framework presented here for how the auditory brain stem encodes the simultaneously occurring acoustic events in speech provides a basis for considering future research and clinical issues. Potential applications range from diagnosing pediatric learning problems to geriatric hearing rehabilitation. Because the auditory brain stem response provides an objective biological index of auditory speech processing, future research along these lines is likely to continue to inform auditory function in both normal and clinical populations.

REFERENCES

Banai, K., Nicol, T., Zecker, S., & Kraus, N. (2005). Brainstem timing: Implications for cortical processing and literacy. *Journal of Neuroscience,* in Press.

Batra, R., Kuwada, S., & Maher, V. L. (1986). The frequency-following response to continuous tones in humans. *Hearing Research*, *21*, 167–177.

Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex (commentary). *Nature Neuroscience*, *3*, 965–966.

Blackburn, C. C., & Sachs, M. B. (1989). Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. *Journal of Neurophysiology*, *62*, 1303–1329.

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

Delgutte, B. (1984). Speech coding in the auditory nerve, II: Processing schemes for vowel-like sounds. *Journal of the Acoustical Society of America*, *75*, 879–886.

Delgutte, B., & Kiang, N. Y. (1984a). Speech coding in the auditory nerve, I: Vowel-like sounds. *Journal of the Acoustical Society of America*, *75*, 866–878.

Delgutte, B, & Kiang, N. Y. (1984b). Speech coding in the auditory nerve, III: Voiceless fricative consonants. *Journal of the Acoustical Society of America*, *75*, 887–896.

Delgutte, B., & Kiang, N. Y. (1984c). Speech coding in the auditory nerve, IV: Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America*, *75*, 897–907.

Delgutte, B., & Kiang, N. Y. (1984d). Speech coding in the auditory nerve, V: Vowels in background noise. *Journal of the Acoustical Society of America*, *75*, 908–918.

Fant, G., (1960). Acoustic theory of speech production. The Hague: Mouton.

Frisina, R. D. (2001). Subcortical neural coding mechanisms for auditory temporal processing. *Hearing Research*, *158*, 1–27.

Frisina, R. D., Smith, R. L., & Chamberlain, S. C. (1990). Encoding of amplitude modulation in the gerbil cochlear nucleus, I: A hierarchy of enhancement. *Hearing Research*, *44*, 99–122.

Galbraith, G. C. (1994). Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalography and Clinical Neurophysiology*, *92*, 321–330.

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, CA: College-Hill Press.

Hayes, E. A., Warrier, C. M., Nicol, T. G., Zecker, S. G, & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clinical Neurophysiology*, *114*, 673–684.

Hood, L. J. (1998). *Clinical applications of the auditory brainstem response*. San Diego, CA: Singular.

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

Johnson, K. L., Nicol, T. G., Zecker, S. G., Wright, B. A., & Kraus, N. (2004). Brainstem timing in learning impaired children with excessive auditory backward masking [Abstract]. *Association for Research in Otolaryngology*, *27*, 118.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, *84*, 541–577.

Kaas, J. H., & Hackett, T. A. (1999). 'What' and 'where' processing in auditory cortex (commentary). *Nature Neuroscience*, *2*, 1045–1047.

Kavanagh, K. T., Harker, L. A, & Tyler, R. S. (1984). Auditory brainstem and middle latency responses. I. Effect of response filtering and waveform identification. II. Threshold responses to a 500-HZ tone pip. *Annals of Otology, Rhinology, and Laryngology Supplement*, *108*, 1–12.

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.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Kraus, N., & Nicol, T. G. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences*, *28*, 176–181.

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

Krishnan, A., Xu, Y., 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. (1997). Neural processing and representation of periodicity pitch. *Acta Oto-laryngolica Supplementum*, *532*, 68–76.

Langner, G., Sams, M., Heil, P., & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, *181*, 665–676.

Langner, G, & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat, I: Neuronal mechanisms. *Journal of Neurophysiology*, *60*, 1799–1822.

Marsh, J. T., & Worden, F. G. (1968). Sound evoked frequency-following responses in the central auditory pathway. *Laryngoscope*, *78*, 1149–1163.

Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Journal of the Acoustical Society of America*, *101*, 1585–1599.

McGee, T., Kraus, N., King, C., Nicol, T., & Carrell, T. D. (1996). Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *Journal of the Acoustical Society of America*, *99*, 3606–3614.

Merzenich, M. M., Knight, P. L, & Roth, G. L. (1975). Representation of cochlea within primary auditory cortex in the cat. *Journal of Neurophysiology*, *38*, 231–249.

Mishkin, M, Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, *6*, 414–417.

Møller, A. R., Jannetta, P. J., & Sekhar, L. N. (1988). Contributions from the auditory nerve to the brain-stem auditory evoked potentials (BAEPs): results of intracranial recording in man. *Electroencephalography and Clinical Neurophysiology*, *71*, 198–211.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 11800–11806.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, *2*, 1131–1136.

Russo, N., Nicol, T. G., Musacchia, G., and Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, *115*, 2021–2030.

Russo, N., 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.

Sayers, B. M., Beagley, H. A., & Henshall, W. R. (1974). The mechanism of auditory evoked EEG responses. *Nature*, *247*, 481–483.

Schreiner, C. E., & Langner, G. (1988). Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *Journal of Neurophysiology*, *60*, 1823–1840.

Sininger, Y., & Starr, A. (2001). *Auditory neuropathy: A new perspective on hearing disorders*. San Diego, CA: Singular.

Sohmer, H., & Pratt, H. (1977). Identification and separation of acoustic frequency following responses (FFRS) in man. *Electroencephalography and Clinical Neurophysiology*, *42*, 493–500.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G., Jr. (1993) Temporal encoding of phonetic features in auditory cortex. *Ann N Y Acad Sci*, *682*, 415–417.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard, M. A., III. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology*, *82*, 2346–2357.

Suga, N., & O'Neill, W. E. (1979). Neural axis representing target range in the auditory cortex of the mustache bat. *Science*, *206*, 351–353.

Suzuki, T., Kobayashi, K., & Hirabayashi, M. (1983). Frequency composition of auditory middle responses. *British Journal of Audiology*, *17*, 1–4.

Wang, X., & Sachs, M. B. (1994). Neural encoding of single-formant stimuli in the cat, II: Responses of anteroventral cochlear nucleus units. *Journal of Neurophysiology*, *71*, 59–78.

Warrier, C. M., Johnson, K. L., Hayes, E. A., Nicol, T., & Kraus, N. (2004). Learning impaired children exhibit timing deficits and training-related improvements in auditory cortical responses to speech in noise. *Experimental Brain Research*, *157*, 431–441.

Wible, B., Nicol, T., & Kraus, N. (2004a). Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biological Psychology*, *67*, 299–317.

Wible, B., Nicol, T., & Kraus, N. (2004b). Encoding of complex sounds in an animal model: Implications for understanding speech perception in humans. In P. Heil, R. König, E. Budinger, & H. Scheich (Eds.), *The auditory cortex: A synthesis of human and animal research.* pp. 241–254. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Yamamoto, K., Sakabe, N., & Kaiho, I. (1979). Power spectral analysis of auditory evoked response. *Journal of the American Auditory Society*, *5*, 107–111.

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *Journal of the Acoustical Society of America*, *66*, 1381–1403.

AQ: 2