

Stimulus Rate and Subcortical Auditory Processing of Speech

Jennifer Krizman^{a, b} Erika Skoe^a Nina Kraus^{a–d}

^aAuditory Neuroscience Laboratory, Department of Communication Sciences, ^bDepartment of Neurobiology and Physiology, ^cDepartment of Otolaryngology, and ^dNorthwestern University Institute of Neuroscience, Northwestern University, Evanston, Ill., USA

Key Words

Stimulation rate · Auditory brainstem response · Frequency-following response · Subcortical auditory structures · Brainstem

Abstract

Many sounds in the environment, including speech, are temporally dynamic. The auditory brainstem is exquisitely sensitive to temporal features of the incoming acoustic stream, and by varying the speed of presentation of these auditory signals it is possible to investigate the precision with which temporal cues are represented at a subcortical level. Therefore, to determine the effects of stimulation rate on the auditory brainstem response (ABR), we recorded evoked responses to both a click and a consonant-vowel speech syllable (/da/) presented at three rates (15.4, 10.9 and 6.9 Hz). We hypothesized that stimulus rate affects the onset to speech-evoked responses to a greater extent than click-evoked responses and that subcomponents of the speech-ABR are distinctively affected. While the click response was invariant with changes in stimulus rate, timing of the onset response to /da/ varied systematically, increasing in peak latency as presentation rate increased. Contrasts between the click- and speech-evoked onset responses likely reflect acoustic differences, where the speech stimulus onset is more gradual, has more delineated spectral information,

and is more susceptible to backward masking by the subsequent formant transition. The frequency-following response (FFR) was also rate dependent, with response magnitude of the higher frequencies (>400 Hz), but not the frequencies corresponding to the fundamental frequency, diminishing with increasing rate. The selective impact of rate on high-frequency components of the FFR implicates the involvement of distinct underlying neural mechanisms for high- versus low-frequency components of the response. Furthermore, the different rate sensitivities of the speech-evoked onset response and subcomponents of the FFR support the involvement of different neural streams for these two responses. Taken together, these differential effects of rate on the ABR components likely reflect distinct aspects of auditory function such that varying rate of presentation of complex stimuli may be expected to elicit unique patterns of abnormality, depending on the clinical population.

Copyright © 2010 S. Karger AG, Basel

Introduction

Hearing depends on accurate neural encoding and perception of temporal events in auditory signals. The auditory brainstem reflects temporal events with extraordinary precision such that minuscule timing delays are diagnostically significant in the assessment of hearing loss

and neurological function [for review see Hall, 1992; Hood, 1998]. The auditory brainstem response (ABR) is a far-field recording of stimulus-locked synchronous neural events. The human ABR to complex sounds reveals distinct aspects of auditory processing in expert and clinical populations that may reflect differences in the encoding and processing of temporal cues. By manipulating the stimulus presentation rate, the effects of neural fatigue and desynchronization become increasingly evident, helping to reveal minute differences in how temporal cues are processed in various subpopulations. Understanding the effects of stimulation rate on the various facets of brainstem activity evoked by complex sounds is fundamental to our knowledge of hearing and its disorders.

The Click-ABR

The ABR is a far-field recording of stimulus-locked synchronous neural events. Acoustic elements that are transient, rapid, and spectrally broad (e.g. clicks) elicit a characteristic pattern of neural activity. The click-ABR, which provides a reliable and noninvasive method for assessing the integrity of peripheral and subcortical auditory structures, is widely used by clinicians when evaluating hearing and the health of the auditory brainstem and periphery [Hall, 2007; Hood, 1998; Sininger, 1993; Starr and Don, 1988]. Timing delays on the order of fractions of milliseconds are clinically relevant in the diagnoses of hearing loss and brainstem pathologies.

Over the last 40 years, a vast literature has been amassed to describe how the click-evoked response changes for patient factors such as age, sex, extent of hearing loss and different stimulus conditions, including intensity and presentation rate. It is generally accepted that for rates between 2 and 20 Hz [Fowler and Noffsinger, 1983; Hall, 2007] and possibly upwards of 30 Hz [Hood, 1998], the click-ABR is invariant. Rates faster than 30 Hz result in latency delays and, in some cases, amplitude reductions [Don et al., 1977], with later response peaks more greatly affected by rate than earlier peaks [Hood, 1998]. Stimulation rate has been used to enhance differences between normal and pathological auditory function. For example, patients with multiple sclerosis are excessively affected by rate [Jacobson et al., 1987] and children with specific language impairments demonstrate greater increases in peak latency to increasing click rates relative to normal-learning children [Basu et al., 2009].

The Frequency-Following Response

ABRs are also elicited by continuous or periodic sounds, such as sinusoidal tones. Brainstem neurons

phase lock to the temporal structure of the eliciting sound, giving rise to a sustained response known as the frequency-following response (FFR), which reflects the encoding of the periodic (i.e. frequency-specific) information of the stimulus ≤ 2 kHz [Hoormann et al., 1992; Moushegian et al., 1973]. The transient (e.g. click-ABR) and sustained ABRs are assumed to originate from separate neural generators [for a review, see Chandrasekaran and Kraus, 2009]. For example, FFR latency and click-ABR latency are not correlated [Hoormann et al., 1992] and are differentially affected by intensity [Akhoun et al., 2008] as well as sex [Hoormann et al., 1992].

The Speech-ABR

In addition to representing the transient features of speech sounds, the auditory brainstem represents steady-state and time-varying formant information. By phase-locking to the fundamental frequency (F_0) and formant-related harmonics of the stimulus, subcortical synchrony is observed in response to synthesized and natural English vowels [Aiken and Picton, 2008; Dajani et al., 2005; Krishnan, 2002], consonant-vowel formant transitions [Akhoun et al., 2008; Banai et al., 2009; Plyler and Ananthanarayan, 2001; Russo et al., 2004], speech syllables [Hornickel et al., 2009b], and words [Galbraith et al., 1995, 1997, 2004]. In fact, ABRs recorded to speech reflect the acoustics with such accuracy that when the evoked response is played back as an auditory stimulus, it is perceived as intelligible speech [Galbraith et al., 1995]. In addition to this fidelity, ABRs are also influenced by lifelong auditory experience with language [Krishnan and Plack, 2009; Krishnan et al., 2005; Swaminathan et al., 2008; for review see Skoe and Kraus, 2010] and music [Kraus et al., 2009; Lee et al., 2009; Musacchia et al., 2007; Parbery-Clark et al., 2009; Strait et al., 2009; Wong et al., 2007; for review see Skoe and Kraus, 2010]. For example, the subcortical response is larger in amplitude to forward as compared to backward speech, suggesting that the brainstem may respond preferentially to familiar sounds [Galbraith et al., 2004]. This experience-dependent plasticity and the link between subcortical processes and higher-level (i.e. cortical) function likely involve top-down modulation of subcortical structures via corticofugal pathways [reviewed in Tzounopoulos and Kraus, 2009].

The evoked brainstem potential in response to a stop consonant speech syllable such as /da/ consists of a transient response similar to the click-ABR [Song et al., 2006], reflecting the transient stop burst of the consonant /d/, and an FFR to the voiced formant transition from the /d/ to the vowel /a/. Stop consonants are especially vulner-

able to misperception in clinical populations, including poor readers [de Gelder and Vroomen, 1998; Tallal, 1980, 1981], people with hearing loss [Townsend and Schwartz, 1981; Van Tasell et al., 1982] and people with auditory-processing disorders [Banai and Kraus, 2008; Bellis, 2002; Tobey et al., 1979]. Considerable work has been done to investigate how the brainstem responds to the speech syllable /da/ [Banai et al., 2005, 2009; Chandrasekaran et al., 2009b; Cunningham et al., 2001; Dhar et al., 2009; Hornickel et al., 2009b; Johnson et al., 2007, 2008; King et al., 2002; Kraus and Nicol, 2005; Russo et al., 2004; Song et al., 2008; Wible et al., 2004; for review see Skoe and Kraus, 2010]. This work has led to the development of BioMARK (Biological Marker of Auditory Processing, Natus Medical Inc.). This clinical technology uses a 40-ms speech syllable /da/ with a standard presentation rate of 10.9 Hz. Like the click-ABR, the speech-ABR to /da/ evokes a characteristic response (fig. 1) which mimics the evoking stimulus.

Akin to the BioMark /da/, the ABR elicited by the speech syllable /da/ used in this study consists of nine characteristic peaks (fig. 1a). Auditory brainstem encoding of stimulus timing is reflected in the latency of the peaks. Peaks I, III, V, and A represent the stimulus onset and are analogous to the click-evoked peaks I, III, V and V_n . Peak C signals the transition from the aperiodic stop burst to the periodic (voiced) formant transition, peaks D, E and F represent the F_0 of the speech sound and O occurs in response to the offset of the stimulus. Neural phase locking to the F_0 is measured in the spectral domain as the spectral peak occurring around 100 Hz and in the time domain by the interpeak intervals (i.e. period) of the prominent periodic peaks of the FFR, namely D, E and F. Between these larger-amplitude pitch-related peaks are smaller-voltage fluctuations which represent the higher-frequency information within the phase-locking capabilities of the brainstem [<2 kHz; Krishnan, 2007; Liu et al., 2006]. This encoding includes the first formant (F_1) range (220–720 Hz) of /da/.

The aims of this study were to investigate interactions between auditory temporal processing and stimulus complexity by examining the effects of stimulus rate on speech- and click-evoked ABRs. We hypothesized that variation in presentation rate has a greater effect on the onset encoding of /da/ relative to the click stimulus due to differences in acoustic complexity. Further motivating this hypothesis is that speech-evoked responses are known to be selectively disrupted in clinical populations despite normal click-evoked responses [Banai et al., 2005, 2009; Chandrasekaran et al., 2009b; Hornickel et al.,

2009b; Song et al., 2006]. We further hypothesized that presentation rate selectively affects specific components of the speech-ABR. Specifically, we hypothesized that the slower components associated with pitch (F_0 and lower harmonics) would be rate invariant while faster components reflecting harmonics and onset timing would vary with stimulus rate. Functional dissociation between these slower and faster aspects has been reported in a number of studies [Banai et al., 2009; Johnson et al., 2007; Kraus and Nicol, 2005; Wible et al., 2004], where the higher harmonics and onset timing are diminished or delayed in children with language impairments despite normal F_0 encoding. This hypothesis is consistent with Krishnan [2002], who found that lower and higher harmonics in the FFR were affected differently by presentation rate and with Basu et al. [2009], who showed that rate effects were not equivalent for all peaks of responses to clicks presented above 30 Hz in children with specific language impairment. To test these hypotheses, ABRs were recorded to a click and speech stimulus at three presentation rates: 6.9, 10.9 and 15.4 Hz in young adults.

Methods

Participants

Eighteen adults, 9 female, aged 21–33 years (mean = 26, SD = 3.48) participated in the study. A full audiogram and a click-evoked ABR at a rate of 31.25 Hz were used to assess normal auditory function at levels peripheral to the brainstem. All individuals had normal audiometric thresholds (≤ 10 dB nHL) from 150 to 8000 Hz and normal click-ABR peak V latencies (5.69 ± 0.18 ms) presented at 45 dB nHL. The click stimulus was also presented across the three presentation rates at this intensity level. Calibration using a sound level meter ensured consistency across stimulus presentation rates throughout the recording session. All procedures were approved by the Institutional Review Board of Northwestern University.

Stimuli and Recording Parameters

Brainstem potentials were elicited by a click stimulus, a 100- μ s square wave with broad spectral content, and the syllable /da/, a 40-ms, five-formant synthesized speech sound [Klatt, 1980] which comprises an initial noise burst and formant transition between the consonant and the vowel. The F_0 and first three formants (F_1 , F_2 , F_3) change linearly over the duration of the stimulus: F_0 from 103 to 125, F_1 from 220 to 720, F_2 from 1700 to 1240 and F_3 from 2580 to 2500 Hz. F_4 and F_5 are constant at 3600 and 4500 Hz, respectively.

Both speech and nonspeech conditions were collected in the same manner within the same recording session using the Biologic Navigator Pro System (Natus Medical Inc., Mundelein, Ill., USA). Responses were differentially recorded from Ag-AgCl electrodes with electrode impedance <5 k Ω , with electrodes placed at Cz (active), the right ear lobe (reference) and forehead (ground).

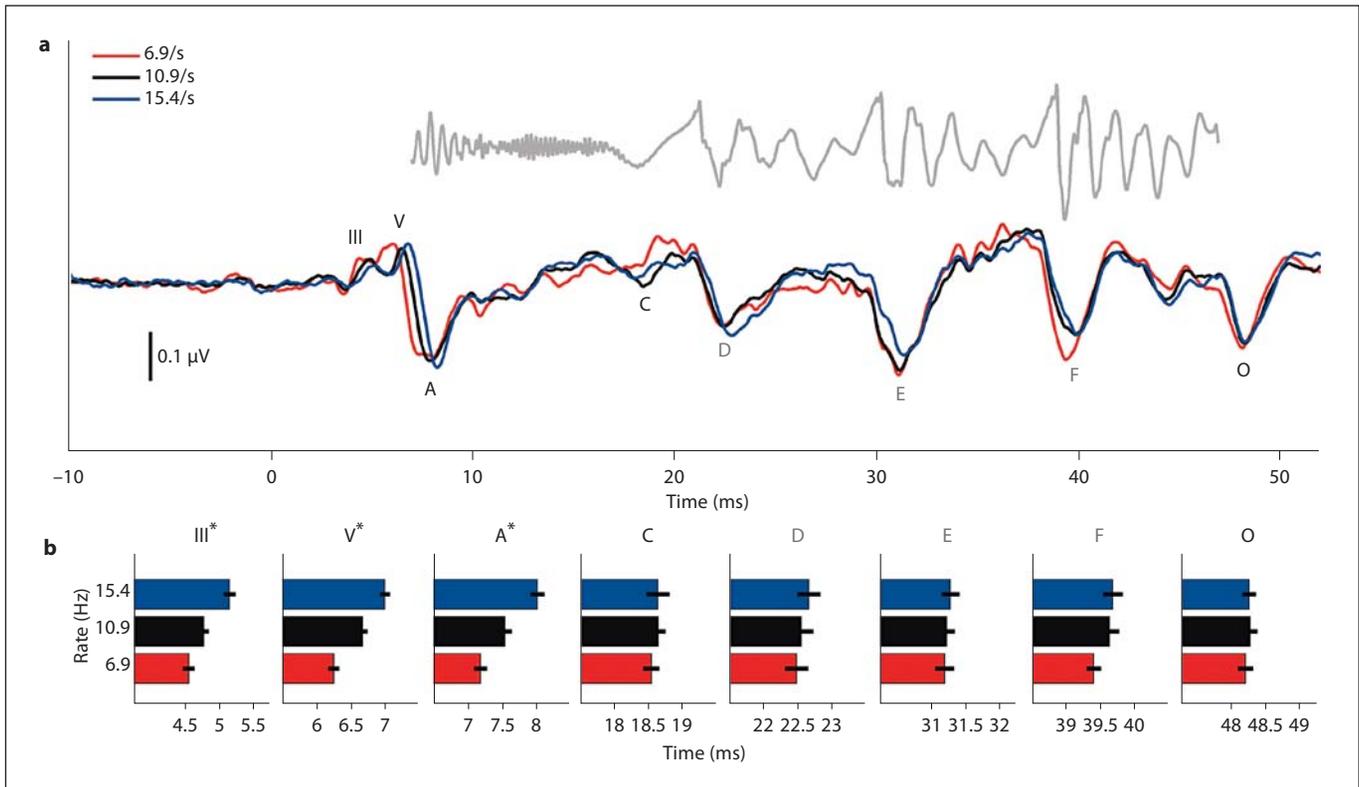


Fig. 1. Comparison of grand average responses and mean peak latencies to /da/ at three stimulation rates. In the onset response to the /da/ stimulus, the peak latencies systematically increased in response to increasing stimulus presentation rates while the subsequent peaks were stable across the three presentation rates. * = Significant results. **a** The grand average responses ($n = 18$) to each rate are plotted with the stimulus plotted in gray above these waveforms. To account for the time associated with neural transmission, the stimulus waveform has been shifted in time to align its onset with the onset of the response for the standard rate. The

peaks D, E, and F, corresponding to the F_0 of the stimulus, are labeled in gray and those corresponding to the timing landmarks are in black (III, V, A, C, and O). **b** Bar graphs showing significant rate effects for all three onset peaks (III, V, and A). In contrast, the peaks of the FFR (D, E, and F) did not shift with rate, with one exception (see text). Additionally, the increased fidelity of higher-frequency aspects of the stimulus at the slower presentation rate is evident in the fine structure of the responses (i.e. better definition of the smaller peaks between D, E and F).

Speech stimuli were presented monaurally to the right ear at 80.3 dB SPL through electromagnetically-shielded insert earphones (ER-3A, Natus Medical Inc.). During testing, each participant watched a DVD of his or her choice with the sound level set to <40 dB SPL, so it could be heard with the unoccluded ear at a level that would not mask the stimulus-evoked response.

For the speech condition, stimuli were presented in alternating polarity and both the click and /da/ were presented at three presentation rates: 15.4 Hz (fast), 10.9 Hz (standard), and 6.9 Hz (slow). Artifact-free ($\pm 23.8 \mu$ V) speech-evoked responses were averaged over a 64-ms time window that included an 11-ms pre-stimulus period to create two subaverages of 3000 sweeps. The click stimulus was presented in a single polarity (i.e. rarefaction) in two 2000-sweep blocks averaged over a 10.66-ms window. The presentation order for the click and speech stimuli always proceeded from fast to slow so as to present the most taxing stimulus condition first. The speech-ABRs were online bandpass filtered from 100 to 2000 Hz (12 dB/octave) and digitally sampled at 16

kHz. The click-ABRs were online bandpass filtered from 100 to 1500 Hz and digitally sampled at 24 kHz. Subaverages were averaged together at the end of the recording session.

Data Analysis

For the click-ABR, peak latency (the time interval between stimulus onset and the peak of the response) and peak amplitude for waves III, V and V_n (the negative trough following V) were visually identified for each subject at each rate. Wave III was identified as the positive peak occurring at approximately 3.8 ms after stimulus onset, wave V was identified as the peak near 5.5 ms immediately before the negative slope, and V_n was selected as the bottom of the downward slope following wave V [Hall, 1992].

The ABR to /da/ has been described in detail and is reliable both within and across subjects [Banai et al., 2005, 2009; Cunningham et al., 2001; Hornickel et al., 2009a; Johnson et al., 2007; King et al., 2002; Russo et al., 2004, 2005; Wible et al., 2004]. For each subject, peak latencies were visually identified and ampli-

Table 1. Percent detection, mean latency, standard deviation and percent of subjects showing the rate effect (i.e. the fast rate is later than the standard rate which is later than the slow rate) at the individual peaks

Peak:	Speech																	
	III			V			A			C			D			E		
Rate, Hz	15.4	10.9	6.9	15.4	10.9	6.9	15.4	10.9	6.9	15.4	10.9	6.9	15.4	10.9	6.9	15.4	10.9	6.9
% Detection	100	100	100	100	100	100	100	100	100	67	67	61	89	94	89	100	100	100
Mean latency, ms	5.16	4.78	4.56	7.02	6.70	6.26	8.04	7.56	7.20	18.64	18.65	18.54	22.73	22.56	22.37	31.29	31.24	31.20
SD, ms	0.37	0.30	0.36	0.31	0.27	0.33	0.42	0.43	0.40	0.72	0.46	0.49	0.73	0.78	0.59	0.55	0.50	0.57
% Showing rate effect	94.40			100			94.40			n/a			27.80			0		

None of the peaks after of the onset response to the speech syllable /da/ had more than 30% of the subjects showing a rate-dependent effect. This pattern was seen in even fewer subjects in the click-evoked response. A possible explanation for the higher number of subjects displaying the pattern in the FFR peaks to /da/ than to the onset peaks to the click is that the strength of the rate dependence on the latency of the /da/ onset peaks affected these later peaks. For example, peak F was the only FFR peak that had a significant effect of rate but was no longer significant when peak A was controlled for in the analyses.

tudes were determined for nine peaks in the ABR, including the onset (I, III, V and A), transition (C), offset (O) and frequency-following (D, E and F) peaks. Peaks I, III, V, and A of the speech-evoked response were picked using similar criteria as for peaks I, III, V, and V_n of the click-evoked response. Peaks C, D, E, F, and O were identified as the deepest troughs within the expected latency range for each peak, consistent with previous reports in young adults [Dhar et al., 2009; Hornickel et al., 2009a]. Average latencies were: C ~18.5 ms, D ~22 ms, E ~31 ms, peak F ~39.7 ms, and the offset peak, O, was centered around 48 ms. Any peak smaller than the amplitude of the prestimulus baseline activity was deemed ‘not reliable’ and excluded from analyses (table 1). Two peaks, I and C, were not analyzed due to their high variability and the difficulty in identifying these peaks in individual subjects. The VA complex was further analyzed by computing the slope, a measure of neural synchrony to the onset of the stimulus. Within the FFR, occurring between 21 and 42 ms (including peaks D, E and F), the average spectral amplitudes of four frequency ranges were analyzed using fast Fourier analysis: F₀, 103–125 Hz, the F₁ frequency range broken into a low and high range, 180–410 and 411–755 Hz, and higher frequencies above the F₁, 756–1130 Hz that are still within the phase-locking capabilities of the brainstem. The F₁ was broken into these two ranges to separate the more prominent frequency peaks in the F₁ response (180–410 Hz) from the less prominent frequencies (411–755 Hz), which pattern with auditory-based learning disabilities [Banai et al., 2009; Johnson et al., 2007].

Repeated-measures analysis of variance (ANOVA) was used to compare the responses to the different presentation rates of the click and speech stimuli. Significance was determined using the Greenhouse-Geisser correction, which determines statistical significance using stricter degrees of freedom. These p values as well as η^2 , a measure of effect size, are reported. For significant F values, Bonferroni post-hoc tests were performed. Data processing were performed using routines coded in MATLAB 2006b (The MathWorks, Inc., Natick, Mass., USA) and statistical analyses were performed in SPSS (SPSS Inc., Chicago, Ill., USA).

Results

Onset Measures Speech-ABR

The onset of the ABR to /da/ was affected by presentation rate. As can be seen in the grand average waveforms in figure 1, the onset response peaks III, V, and A differed significantly in latency such that the faster the presentation rate the later the response. This shift in latency for the onset peaks, evident in the grand average waveforms, is not simply inherited by subsequent peaks. That is, peaks D, E, F and O varied relatively less than the preceding peaks with changing rate. The mean latencies of peaks III, V and A are plotted in figure 1b and the bar graphs illustrate the latency shifts with changes in rate. The pattern of increased latency with increased rate was consistent across subjects, evident in 94.4% of subjects for peaks III and A, and in 100% of the subjects for peak V (table 1). The subject who did not display the pattern at peak III was not the same subject who did not display the pattern at peak A. Peak III latency became systematically later with increasing stimulus rate [$F(1.5, 26.1) = 50.381, p < 0.0005, \eta^2 = 0.748$]. Peak V demonstrated the same pattern [$F(1.96, 33.33) = 286.802, p < 0.0005, \eta^2 = 0.944$], as did peak A [$F(1.92, 32.57) = 213.724, p < 0.0005, \eta^2 = 0.926$]. The amplitude of all onset peaks did not differ across the three rates.

Click-ABR

In line with previous findings, the timing of the click response did not vary across the three rates (fig. 2a). The

Table 1 (continued)

Peak:	Speech						Click								
	F		O		III		V		V _n						
Rate, Hz	15.4	10.9	6.9	15.4	10.9	6.9	15.4	10.9	6.9	15.4	10.9	6.9	15.4	10.9	6.9
% Detection	100	100	100	100	89	94	100	100	100	100	100	100	100	100	100
Mean latency, ms	39.68	39.64	39.41	48.25	48.33	48.24	3.82	3.81	3.81	5.58	5.57	5.57	6.25	6.23	6.23
SD, ms	0.61	0.58	0.44	0.42	0.44	0.45	0.13	0.13	0.14	0.19	0.19	0.19	0.21	0.23	0.24
Showing % rate effect	27.80		16.70		11.10		5.56		17						

grand average click-evoked responses to the three rates were nearly identical in peak latencies and amplitudes. There was no significant difference in latency at peak III [F(1.5, 26.3) < 1, n.s.], peak V [F(1.8, 30.2) < 1, n.s.] or peak V_n [F(1.9, 32.8) < 1, n.s.] in response to the three presentation rates (fig. 2b). There was, however, a significant difference in the amplitude of peak III [F(1.8, 30.6) = 6.140, p = 0.007, η² = 0.265]; post-hoc analyses revealed differences between the fast and slow conditions (p = 0.002) but not between the fast and standard conditions (p = 0.210) or the standard and slow conditions (p = 0.618). No effect of rate was seen for the amplitudes of wave V [F(1.9, 31.8) = 1.902, p = 0.168] or V_n [F(1.8, 29.9) < 1, n.s.].

Click- versus Speech-ABR

Figure 2 compares the latency-dependent onset response to /da/ and the rate-invariant click response (peaks III, V and V_n). A 3 (rate) × 2 (stimuli) repeated measures ANOVA was performed to evaluate rate-dependent effects at these three peaks across the two stimuli. Figure 2b shows the latency shifts of the click-evoked and speech-evoked peaks at the fast, standard and slow rate. The interaction between the three rates and two stimuli was significant across all peaks: III [F(1.6, 27.7) = 48.811, p < 0.0005, η² = 0.742], V [F(1.92, 32.7) = 251.536, p < 0.0005, η² = 0.937], and A/V_n [F(1.8, 30.8) = 206.321, p < 0.0005, η² = 0.924]. The rate-latency plots in figure 2b demonstrate the different effects of presentation rate for the speech and nonspeech conditions on the onset latencies.

Table 2. Repeated-measures ANOVA F values for each measure

Measure	F	p value		
		fast vs. standard	standard vs. slow	fast vs. slow
<i>Speech</i>				
Latency				
III	50.381	<0.0005	0.024	<0.0005
V	286.802	<0.0005	<0.0005	<0.0005
A	213.724	<0.0005	<0.0005	<0.0005
D	<1	n.s.	n.s.	n.s.
E	<1	n.s.	n.s.	n.s.
F	13.843	0.005	n.s.	0.002
O	<1	n.s.	n.s.	n.s.
IPL				
D to E	<1	n.s.	n.s.	n.s.
E to F	5.879	n.s.	0.031	n.s.
FFT				
F ₀	<1	n.s.	n.s.	n.s.
Low F ₁	1.234	n.s.	n.s.	n.s.
High F ₁	10.966	n.s.	0.003	0.009
HF	46.122	0.004	<0.0005	<0.0005
<i>Click</i>				
Latency				
III	<1	n.s.	n.s.	n.s.
V	<1	n.s.	n.s.	n.s.
V _n	<1	n.s.	n.s.	n.s.

Where the repeated measures ANOVA is significant (see text), the Bonferroni-corrected post-hoc p values are given for group comparisons. The greatest difference between conditions was seen in the speech onset peaks and high harmonics of the FFR response. HF = High frequency.

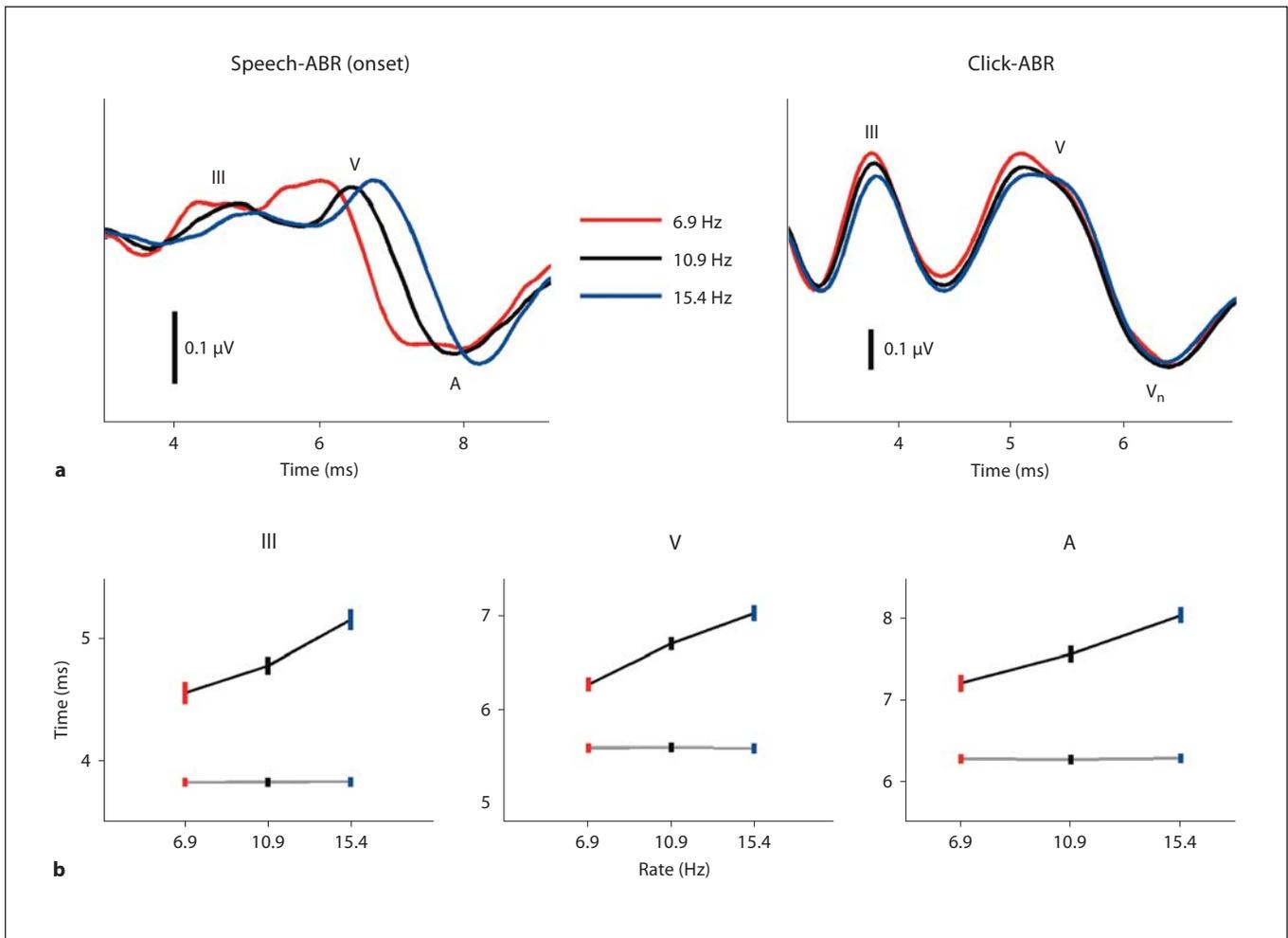


Fig. 2. Comparison of analogous peaks of the speech- and click-evoked responses at the three presentation rates. **a** Peaks III, V, and A/V_n of the click- and speech-ABRs are plotted for each stimulus presentation rate. **b** Rate-latency functions for peaks III, V, and A/V_n. Both click-evoked (gray line) and speech-evoked (black

line) functions are plotted. The height of the bar represents standard error of the peak latency. For speech, the latency of all three onset peaks shifts progressively later with increasing rate. The peak latencies of III, V, and V_n of the click-evoked response do not shift with rate.

FFR of the Speech-ABR Fundamental Frequency

As hypothesized, rate affected the timing of the onset of the speech-ABR but had little effect on the timing of the FFR peaks D, E and F, which reflect the subcortical encoding of the F₀ (fig. 1). Although there was an effect of rate at peak F [$F(1.4, 24.4) = 13.843, p < 0.0005, \eta^2 = 0.449$], the effect was not present for all pairwise comparisons (table 2). When the latencies at peak F were covaried with the latencies at peak A, the rate effect disappeared [$F(1.4, 19.8) < 1, n.s.$], suggesting that the shift seen at peak F is a carryover of the large effect of rate on the onset response.

To further examine the subcortical encoding of the F₀, the spectra of the responses from 21 to 42 ms across the three presentation rates were analyzed (fig. 3). As in the temporal domain, mean spectral amplitude of the F₀ range (103–125 Hz) was invariant across the three rates [$F(1.9, 32.3) < 1, n.s.$]. Consistent with these findings, the difference in interpeak latency, of D to E, reflecting the period of the F₀, did not differ with rate [$F(1.4, 23.9) < 1, n.s.$]. Peak E to peak F interpeak latency did show a significant effect of rate [$F(1.8, 30.1) = 5.879, p = 0.006, \eta^2 = 0.257$], although this was only significant between the responses to the standard and slow presentation rates. Lastly, the offset peak, O, showed no effect of rate

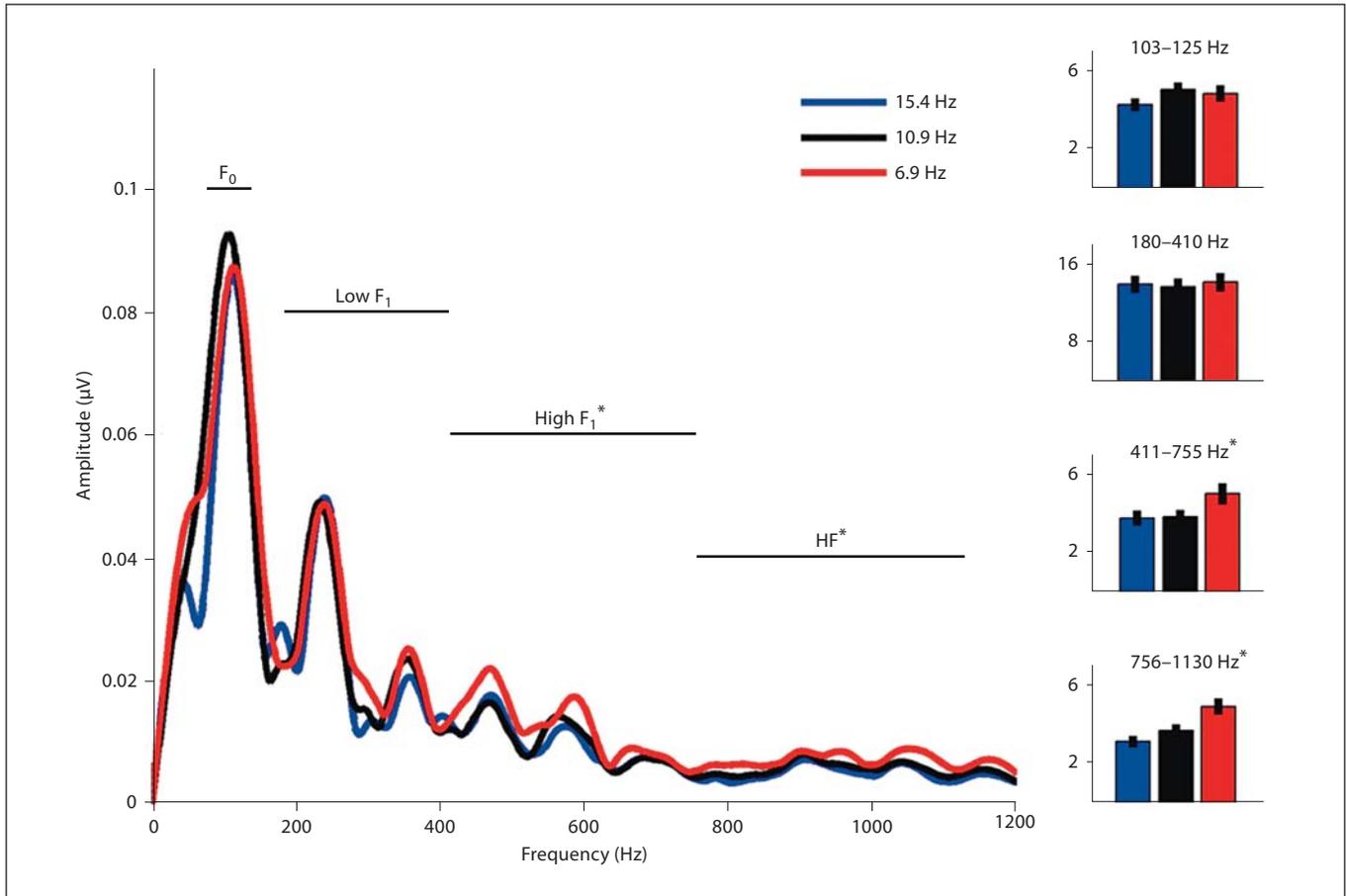


Fig. 3. Frequency spectra of the speech-evoked FFR region for three presentation rates. For the bar graphs, the y-axis of the top graph is multiplied by 10^{-2} and the remaining y-axes are multiplied by 10^{-3} . A significant effect of rate (*) was seen in both the

high F₁ (411–755 Hz) and high-frequency (HF, 711–1130 Hz) range (see text). There were no significant effects of rate on frequencies centered around the F₀ (103–125 Hz) or low F₁ (180–410 Hz).

[$F(1.42, 24.1) < 1$, n.s.]. Therefore, the findings in the spectral domain complement the findings in the temporal domain, confirming that F₀ encoding is stable with rate.

Harmonics

Similar to the F₀, the spectral amplitude of the lower region of F₁, from 180 Hz to 410 Hz, did not vary with the presentation rate [$F(1.5, 25.9) = 1.234$, $p = 0.298$, n.s.]. The encoding of the high range of F₁ from 411 to 755 Hz, however, was rate dependent [$F(1.6, 26.7) = 10.966$, $p = 0.001$, $\eta^2 = 0.392$]. In this range, the difference was significant between the responses to the fast and slow stimulus presentations (table 2) as well as between the standard and slow responses but not between the fast to standard stimulus presentation rates. Additionally, the

average spectral amplitude of the higher harmonics from 756 to 1130 Hz increased as the presentation rate slowed [$F(1.6, 26.5) = 46.122$, $p < 0.0005$, $\eta^2 = 0.731$] (fig. 3). This monotonic increase was significant across all stimulus presentation rates.

Discussion

The effects of stimulus timing on the human ABR depend on the acoustics of the evoking stimulus and the aspect of brainstem activity considered. Rate had a dramatic affect on the timing of the onset portion of the speech-evoked response while corresponding click-evoked peaks were invariant. Rate affected the FFR in a systematic manner, with higher frequencies becoming

increasingly rate sensitive while lower frequencies (notably the F_0) remained rate resistant.

Onset response differences between speech and click stimuli can be attributed to stimulus differences. Whereas clicks contain a broad range of frequencies, speech is more spectrally shaped. In addition, the onset of the /da/ stimulus occurs more gradually relative to the instantaneous rise time of the click. The onset of the /da/ syllable may also be more susceptible to the effects of backward masking by the larger-amplitude formant transition [Johnson et al., 2007]. Finally, brainstem activity can be experience dependent [Tzounopoulos and Kraus, 2009], i.e. the differing rate effects of the two stimuli may be due to the greater exposure to and use of speech sounds.

Another consideration when interpreting the rate effects for the speech versus click stimuli is that although the presentation rates were identical, the click is shorter in duration resulting in a longer interstimulus interval (ISI) relative to the speech stimulus. For the presentation rates used here, the ISIs were 145 versus 105 ms (slow), 92 versus 52 ms (standard), and 65 versus 25 ms (fast) for the click versus speech stimuli, respectively. In order for the /da/ and click stimuli to occur at equivalent ISIs, the rates for the click stimulus would need to be 40, 19.2 and 9.5 Hz to obtain ISIs equivalent to the /da/ at the fast, standard and slow rates, respectively. For two of these presentation rates, 9.5 and 19.2 Hz, click-evoked response latencies are known to be rate invariant [Fowler and Noffsinger, 1983; Hall, 1992; Hood, 1998]. Thus, the differences observed between the onset response of the speech and click-ABRs cannot be accounted for by differences in ISI. These results suggest that the encoding of certain sounds is more resistant to the stress of increased stimulation rate than other sounds.

The effect of stimulus presentation rate is likely bounded by a maximum and minimum rate, where rates outside either extreme would no longer affect the response, and rates near the extremes would show nonlinear outcomes. These boundaries likely reflect an interaction of neural adaptation, neural fatigue, and refractory properties of individual nerve fibers resulting in a desynchronization of the response that most affects the encoding of the faster elements of the stimulus [Hall, 1992; Jacobson et al., 1987]. Varying the presentation rate, then, manipulates the neurophysiological mechanisms underlying the subcortical encoding of timing, thereby elucidating what happens to the population-wide neural response when the stimulus is manipulated along this temporal dimension.

For the speech stimulus, the onset response and the lower frequency components of the FFR were affected

differently by stimulation rate, suggesting that these response components enlist distinct neural populations in the auditory pathway. Considerable data support the existence of separate neural mechanisms for the onset response and FFR [Akhoun et al., 2008; Chandrasekaran and Kraus, 2009a; Hoormann et al., 1992; Hornickel et al., 2009b]. Using a longer speech syllable (a 60-ms /ba/), Akhoun et al. [2008] found that as stimulus intensity decreased, the onset response and FFR both increased in latency. However, the FFR increased at a greater rate than the onset response. Background noise is also known to diminish the onset response while the FFR continues to be robust [Russo et al., 2004]. Thus, stimulus manipulations have different impacts on the onset response and FFR. Moreover, compared to the FFR, the transient onset is less susceptible to changes associated with short-term auditory training [Russo et al., 2005].

Stressing the system degrades ABRs even in the normal auditory system [Galbraith et al., 2004, 1995; Russo et al., 2004; Song et al., 2006]. This degradation is inordinately exacerbated in clinical populations when the stimulus is more ecologically valid [Banai et al., 2009; Chandrasekaran et al., 2009b; Hornickel et al., 2009a; Wible et al., 2004], presented in background noise [Russo et al., 2004] or at a faster rate [Basu et al., 2009]. Thus, impairments are feature-specific and not generalized or pan-response phenomena. Children with reading impairments, for example, have normal click-ABRs and normal F_0 encoding, yet abnormal responses to the faster elements of speech (i.e. harmonics and timing) [Banai et al., 2009; Cunningham et al., 2001]. Furthermore, long-term music and language experience selectively enhance specific stimulus features of brainstem activity [Krishnan et al., 2005, 2009; Lee et al., 2009; Musacchia et al., 2007; Strait et al., 2009; Swaminathan et al., 2008; Wong et al., 2007]. Consistent with this previous work, rate effects do not generalize to the entire response but are specific to the onset and higher-frequency subcomponents of brainstem activity, primarily in response to the faster elements of speech.

Stimulus rate disproportionately affects subcomponents of human brainstem activity, specifically the faster elements of speech, and thereby provides an index for examining the role of subcortical timing and its relationship to normal, impaired and expert auditory perception. The rate effects demonstrated here in normal-hearing young adults are likely to be more pronounced in populations where auditory processing is compromised, such as older adults or reading-impaired children, who have decreased neural synchrony and impaired perception of rapid speech elements [Casparly et al., 1995; Gordon-

Salant et al., 2007; Merzenich et al., 1996; Tallal et al., 1985]. Varying stimulus presentation rate, then, is expected to have different neural consequences in expert, normal and impaired populations. Further investigation into the effects of stimulus rate will continue to reveal the interplay between stimulus timing and temporal processing, its role in perception, and the underlying mechanisms that are selectively enhanced or diminished in expert and clinical populations.

Acknowledgements

The authors thank Trent Nicol, Catherine Warrier, Dana Strait and other members of the Auditory Neuroscience Laboratory for their comments on the manuscript. This work was supported by NIH R01 DC01510, F32 DC008052.

References

- Aiken SJ, Picton TW: Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 2008;245:35–47.
- Akhoun I, Gallego S, Moulin A, Menard M, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H: The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol* 2008;119:922–933.
- Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N: Reading and subcortical auditory function. *Cereb Cortex* 2009;19:2699–2707.
- Banai K, Kraus N: The dynamic brainstem: Implications for APD; in McFarland D, Cacace A (eds): *Current Controversies in Central Auditory Processing Disorder*. San Diego, Plural Publishing, 2008, pp 269–289.
- Banai K, Nicol T, Zecker SG, Kraus N: Brainstem timing: implications for cortical processing and literacy. *J Neurosci* 2005;25:9850–9857.
- Basu M, Krishnan A, Weber-Fox C: Brainstem correlates of temporal auditory processing in children with specific language impairment. *Develop Sci* 2009;13:77–91.
- Bellis TJ: *When the Brain Can't Hear: Unraveling the Mystery of Auditory Processing Disorder*. New York, Pocket Books, 2002.
- Caspary DM, Milbrandt JC, Helfert RH: Central auditory aging: GABA changes in the inferior colliculus. *Exp Gerontol* 1995;30:349–360.
- Chandrasekaran B, Kraus N: The scalp-recorded brainstem response to speech: neural origins. *Psychophysiology* 2009a; doi: [10.1111/j.1469-8986.2009.00928.x](https://doi.org/10.1111/j.1469-8986.2009.00928.x).
- Chandrasekaran B, Hornickel J, Skoe E, Nicol T, Kraus N: Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron* 2009b; 64: 311–319.
- Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N: Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clin Neurophysiol* 2001;112:758–767.
- Dajani HR, Purcell D, Willy W, Kunov H, Picton TW: Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Trans Biomed Eng* 2005;52: 1614–1618.
- de Gelder B, Vroomen J: Impaired speech perception in poor readers: Evidence from hearing and speech reading. *Brain Lang* 1998;64: 269–281.
- Dhar S, Abel R, Hornickel J, Nicol T, Skoe E, Zhao W, Kraus N: Exploring the relationship between physiological measures of cochlear and brainstem function. *Clin Neurophysiol* 2009;120:959–966.
- Don M, Allen AR, Starr A: Effect of click rate on the latency of auditory brain stem responses in humans. *Ann Otol Rhinol Laryngol* 1977; 86:186–195.
- Fowler CG, Noffsinger D: Effects of stimulus repetition rate and frequency on the auditory brainstem response in normal cochlear-impaired, and VIII nerve/brainstem-impaired subjects. *J Speech Hear Res* 1983;26:560–567.
- Galbraith GC, Amaya EM, de Rivera JM, Donan NM, Duong MT, Hsu JN, Tran K, Tsang LP: Brain stem evoked response to forward and reversed speech in humans. *Neuroreport* 2004;15:2057–2060.
- Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM: Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 1995;6:2363–2367.
- Galbraith GC, Jhaveri SP, Kuo J: Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition. *Electroencephalogr Clin Neurophysiol* 1997;102:46–53.
- Gordon-Salant S, Fitzgibbons PJ, Friedman SA: Recognition of time-compressed and natural speech with selective temporal enhancements by young and elderly listeners. *J Speech Lang Hear Res* 2007;50:1181–1193.
- Hall JW: *Handbook of Auditory Evoked Responses*. Boston, Allyn & Bacon, 1992.
- Hall JW: *New Handbook of Auditory Evoked Responses*. Boston, Pearson, 2007.
- Hood LJ: *Clinical applications of the auditory brainstem response*. San Diego, Singular Pub Group, 1998.
- Hoormann J, Falkenstein M, Hohnsbein J, Blanke L: The human frequency-following response (FFR): Normal variability and relation to the click-evoked brainstem response. *Hear Res* 1992;59:179–188.
- Hornickel J, Skoe E, Kraus N: Subcortical laterality of speech encoding. *Audiol Neurotol* 2009a;14:198–207.
- Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N: Subcortical differentiation of voiced stop consonants: Relationships to reading and speech in noise perception. *Proc Natl Acad Sci USA* 2009b;106:13022–13027.
- Jacobson JT, Murray TJ, Deppe U: The effects of ABR stimulus repetition rate in multiple sclerosis. *Ear Hear* 1987;8:115–120.
- Johnson KL, Nicol T, Zecker SG, Kraus N: Developmental plasticity in the human auditory brainstem. *J Neurosci* 2008;28:4000–4007.
- Johnson KL, Nicol TG, Zecker SG, Kraus N: Auditory brainstem correlates of perceptual timing deficits. *J Cogn Neurosci* 2007;19: 376–385.
- King C, Warrier CM, Hayes E, Kraus N: Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett* 2002;319:111–115.
- Klatt D: Software for cascade/parallel formant synthesizer. *J Acoust Soc Am* 1980;67:971–975.
- Kraus N, Nicol T: Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci* 2005;28:176–181.
- Kraus N, Skoe E, Parbery-Clark A, Ashley R: Experience-induced malleability in neural encoding of pitch, timbre and timing: implications for language and music. *Ann NY Acad Sci* 2009;1169:543–557.
- Krishnan A: Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 2002;166:192–201.
- Krishnan A: Frequency-following response; in Burkard RF, Eggermont JJ, Don M (eds): *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Philadelphia, Lippincott, Williams & Wilkins, 2007, pp 313–334.

- Krishnan A, Gandour JT, Bidelman GM, Swaminathan J: Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport* 2009;20:408–413.
- Krishnan A, Plack CJ: Auditory brainstem correlates of basilar membrane nonlinearity in humans. *Audiol Neurootol* 2009;14:88–97.
- Krishnan A, Xu Y, Gandour J, Cariani P: Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn Brain Res* 2005;25:161–168.
- Lee KM, Skoe E, Kraus N, Ashley R: Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 2009;29:5832–5840.
- Liu LF, Palmer AR, Wallace MN: Phase-locked responses to pure tones in the inferior colliculus. *J Neurophysiol* 2006;95:1926–1935.
- Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P: Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* 1996;271:77–81.
- Moushegian G, Rupert AL, Stillman RD: Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr Clin Neurophysiol* 1973;35:665–667.
- Musacchia G, Sams M, Skoe E, Kraus N: Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 2007;104:15894–15898.
- Parbery-Clark A, Skoe E, Kraus N: Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci* 2009;29:14100–14107.
- Plyler PN, Ananthanarayan AK: Human frequency-following responses: representation of second formant transitions in normal-hearing and hearing-impaired listeners. *J Am Acad Audiol* 2001;12:523–533.
- Russo N, Nicol T, Musacchia G, Kraus N: Brainstem responses to speech syllables. *Clin Neurophysiol* 2004;115:2021–2030.
- Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N: Auditory training improves neural timing in the human brainstem. *Behav Brain Res* 2005;156:95–103.
- Sininger YS: Auditory brain stem response for objective measures of hearing. *Ear Hear* 1993;14:23–30.
- Skoe E, Kraus N: Auditory brainstem response to complex sounds: a tutorial. *Ear Hear* DOI: [10.1097/AUD.0b013e3181c9db272](https://doi.org/10.1097/AUD.0b013e3181c9db272).
- Song JH, Banai K, Kraus N: Brainstem timing deficits in children with learning impairment may result from corticofugal origins. *Audiol Neurootol* 2008;13:335–344.
- Song JH, Banai K, Russo NM, Kraus N: On the relationship between speech- and non-speech-evoked auditory brainstem responses. *Audiol Neurootol* 2006;11:233–241.
- Starr A, Don M: Brain potentials evoked by acoustic stimuli; in Picton TW (ed): *Handbook of Electroencephalography and Clinical Neurophysiology*. Amsterdam, Elsevier, 1988, pp 97–150.
- Strait DL, Kraus N, Skoe E, Ashley R: Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci* 2009;29:661–668.
- Swaminathan J, Krishnan A, Gandour JT: Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 2008;19:1163–1167.
- Tallal P: Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang* 1980;9:182–198.
- Tallal P: Language disabilities in children: Perceptual correlates. *Int J Pediatr Otorhinolaryngol* 1981;3:1–13.
- Tallal P, Stark RE, Mellits D: The relationship between auditory temporal analysis and receptive language development: evidence from studies of developmental language disorder. *Neuropsychologia* 1985;23:527–534.
- Tobey EA, Cullen JK Jr, Rampp DL: Effects of stimulus-onset asynchrony on the dichotic performance of children with auditory-processing disorders. *J Speech Hear Res* 1979;22:197–211.
- Townsend TH, Schwartz DM: Error analysis on the California Consonant Test by manner of articulation. *Ear Hear* 1981;2:108–111.
- Tzounopoulos T, Kraus N: Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 2009;62:463–469.
- Van Tasell DJ, Hagen LT, Koblas LL, Penner SG: Perception of short-term spectral cues for stop consonant place by normal and hearing-impaired subjects. *J Acoust Soc Am* 1982;72:1771–1780.
- Wible B, Nicol T, Kraus N: Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* 2004;67:299–317.
- Wong PCM, Skoe E, Russo NM, Dees T, Kraus N: Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 2007;10:420–422.