Research Paper

Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children

Jane Hornickel [a,*], Erica Knowles [a], Nina Kraus [a,b,c]

[a] Auditory Neuroscience Laboratory, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA
[b] Department of Neurobiology and Physiology, Northwestern University, 633 Clark Street, Evanston, IL 60208, USA
[c] Department of Otolaryngology, Northwestern University, 633 Clark Street, Evanston, IL 60208, USA

ABSTRACT

The click-evoked auditory brainstem response (ABR) is widely used in clinical settings, partly due to its predictability and high test-retest consistency. More recently, the speech-evoked ABR has been used to evaluate subcortical processing of complex signals, allowing for the objective assessment of biological processes underlying auditory function and auditory processing deficits not revealed by responses to clicks. Test-retest reliability of some components of speech-evoked ABRs has been shown for adults and children over the course of months. However, a systematic study of the consistency of the speech-evoked brainstem response in school-age children has not been conducted. In the present study, speech-evoked ABRs were collected from 26 typically-developing children (ages 8–13) at two time points separated by one year. ABRs were collected for /da/ presented in quiet and in a 6-talker babble background noise. Test-retest consistency of response timing, spectral encoding, and signal-to-noise ratio was assessed. Response timing and spectral encoding were highly replicable over the course of one year. The consistency of response timing and spectral encoding found for the speech-evoked ABRs of typically-developing children suggests that the speech-evoked ABR may be a unique tool for research and clinical assessment of auditory function, particularly with respect to auditory-based communication skills.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The auditory brainstem response (ABR) is a far-field measure that reflects the synchronous neural activity generated by nuclei along the brainstem in response to an acoustic signal. For many reasons, including the high replicability of the response, the click-evoked ABR has been widely used as a clinical measure of auditory thresholds and neural integrity (Hall, 2006; Hood, 1998; Sininger, 2007). Within an individual, the variability in response timing (peak latencies) from one test to another is small, even when the tests are conducted months apart (Edwards et al., 1982; Lauter and Oyler, 1992; Tusa et al., 1994). Response variability for both adults and children is higher across individuals than within individuals (Edwards et al., 1982; Lauter and Oyler, 1992). The replicability and precise temporal resolution of the click-evoked ABR is such that changes in the absolute latencies from test to retest on the order of fractions of a millisecond are considered clinically

significant (Hall, 2006). In children, the click-evoked response is mature and adult-like by the age of two years (Hood, 1998; Ponton et al., 2000; Salamy, 1984; Sininger, 2007), although there are additional maturational changes that occur throughout childhood such as decreases in response amplitude and response variability (Issa and Ross, 1995; Lauter and Oyler, 1992; Ponton et al., 2000; Salamy, 1984).

ABRs can also be elicited by more complex stimuli such as speech or music. Similar to the click-evoked ABR, the speech-evoked ABR is highly replicable in young adults (Song et al., 2011a). Because the speech-evoked ABR appears to be mature by the age of five (Johnson et al., 2008), it is also likely to be replicable in school-age children. ABRs to more complex stimuli, such as speech or music, retain both the temporal and spectral characteristics of the evoking sound with remarkable fidelity (Russo et al., 2004; Skoe and Kraus, 2010), requiring precise neural phase-locking and synchronous responses to transient stimuli.

While the test-retest reliability of speech-ABR measures has been established in adults (Song et al., 2011a), a systematic study of the test-retest reliability of speech-evoked brainstem responses in typically-developing children has not been conducted. Some response measures appear to be stable in children with a range of

* Corresponding author. Tel.: +1 847 491 2459; fax: +1 847 491 4975.
E-mail address: j-hornickel@northwestern.edu (J. Hornickel).
URL: http://www.brainvolts.northwestern.edu

neurodevelopmental disorders (Russo et al., 2005, 2010), however the test-retest interval was short, on average five months, and response components assessed did not include measures of response timing and spectral encoding recently shown to relate to auditory-based communication skills (Anderson et al., 2010a,b; Banai et al., 2009; Hornickel et al., 2009, 2011; Wible et al., 2004). If the speech-evoked ABR is to be used to investigate and clinically evaluate auditory function and possibly serve as an objective metric of experience-dependent change, the reliability of speech-evoked ABR measures needs to be systematically assessed in school-age children.

The present study investigates the test-retest reliability of speech-evoked auditory brainstem response measures of timing, spectral encoding, and within-session consistency in typically-developing children. Importantly, these measures, particularly response timing, have been linked to auditory communication skills such as reading and listening to speech in noisy environments (Anderson et al., 2010a,b; Banai et al., 2009; Hornickel et al., 2009, 2011; Wible et al., 2004). Given that both the click- and speech-evoked ABRs are thought to be mature by five years of age (Johnson et al., 2008; Salamy, 1984), we do not expect to see a change in brainstem response properties over a 12 month time interval in typically-developing children ages 8–13 years old. By demonstrating the consistency of these measures, we can establish the stability of the speech-evoked brainstem response and provide norms for test-retest change in typically-developing children which can be used to assess auditory function and evaluate change in groups undergoing training or receiving intervention.

2. Materials and methods

2.1. Participants

Participants were 26 typically-developing children, ages 8–13 years (mean = 10.5, 12 girls). All children were recruited from the Chicago area. All participants had normal hearing defined as air conduction thresholds <20 dB HL for octaves from 250 to 8000 Hz with air-bone threshold gaps <10 dB for octaves 500–4000 Hz, normal click-evoked brainstem responses (100 μs stimulus presented at 31.3 Hz and 80 dB SPL), full-scale IQ scores ≥85 on the Wechsler Abbreviated Scale of Intelligence (Woerner and Overstreet, 1999), no current or prior neurological disorders, no family history of learning impairments, scores ≥95 on the Test of Word Reading Efficiency (Torgensen et al., 1999), and were not receiving special services in school. The participants were tested during two sessions (Year 1 and Year 2) that occurred on an average of 11.9 months apart (SD = 1.85). All procedures were approved by the Northwestern University Institutional Review Board and complied with the Declaration of Helsinki, and the children and their parents gave informed assent and consent, respectively. Due to the corruption of data files, one participant's response in background noise from Year 1 was lost. This participant was excluded from the test-retest analyses of the response measures for /da/ presented in noise.

2.2. Electrophysiological stimuli and recording parameters

The six-formant speech stimulus /da/ was 170 ms in length (50 ms formant transition and 120 ms steady-state vowel) with a stable fundamental frequency (100 Hz) and fourth (3300 Hz), fifth (3750 Hz), and sixth (4900 Hz) formants. During the formant transition period, the first, second, and third formants were dynamic, rising from 400 to 720 Hz, falling from 1700 to 1240 Hz, and falling from 2580 to 2500 Hz, respectively. The /da/ stimulus was presented in quiet and in the presence of six-talker babble background noise. The six-talker babble was made up of four

female and two male voices speaking grammatically correct but nonsensical sentences, created in Cool Edit Pro, Version 2.1 (Syntrillium Software: Syntrillium Software Corporation 2003). The babble track was 4.7 s long with the signal-to-noise ratio set at +10 dB based on the root mean square amplitude of the entire track. The /da/ stimulus was synthesized using Klatt (1980).

Stimuli of alternating polarity were presented with a 60 ms interstimulus interval at 80 dB SPL through an insert earphone (ER-3, Etymotic Research; Elk Grove Village, IL, USA) to the right ear using the stimulus presentation software Neuroscan Stim 2 (Compumedics; Charlotte, NC, USA). Responses were recorded with a vertical electrode montage (active Cz, forehead ground, and ipsilateral earlobe reference) using Neuroscan Acquire 4.3 (Neuroscan Scan, Compumedics; Charlotte, NC, USA). Responses were collected at a sampling rate of 20 kHz. During electrophysiological recording, participants watched a movie of their choice while seated in a comfortable chair. Participants were able to hear the soundtrack of the movie, played at <40 dB SPL in the testing booth, through their unoccluded left ear. Allowing participants to view a movie during testing encouraged participant cooperation in sitting quietly and relaxed for the testing session.

2.3. Data processing and analyses

All data processing and analysis techniques replicated previously published studies (Anderson et al., 2010a; Song et al., 2011a). Responses were bandpass filtered from 70 to 2000 Hz (12 dB/octave roll-off) and broken into 230 ms analysis windows (40 ms of pre-stimulus activity). Trials with amplitude greater than ±35 μV were excluded. Responses to individual polarities were averaged across the first and second halves of the recording independently and also across the entire recording and then added to form two replications of 3000 sweeps (1500 of each polarity) and one final average of 6000 sweeps (3000 of each polarity). The addition of responses to alternating polarities eliminates the cochlear microphonic and reduces the impact of stimulus artifact, which we did not observe as the use of tube-insert earphones, alternating polarities, and common mode referencing effectively minimizes stimulus artifact; (Aiken and Picton, 2008; Campbell et al., 2011; Gorga et al., 1985).

2.3.1. Response timing
2.3.1.1. Response latencies. Response peaks and troughs within the region corresponding to the onset and formant transition of the stimulus (0–60 ms) were manually identified for responses to /da/ presented in quiet and in noise. These included the onset response (p9 and p10) and the peaks and troughs mimicking the glottal pulse of the stimulus occurring approximately every 10 ms (p42, p43, p52, p53). Peak and trough names were determined by their approximate latencies in milliseconds. Response peaks and troughs occurring at approximately 22, 23, 32, and 33 ms were excluded from the analysis due to their limited reliability in the noise responses. Manual identification of the peaks and troughs was conducted by the first author for both Year 1 and Year 2.

2.3.1.2. Phase shift. Replicability of response timing was further evaluated through comparisons of response phase. Using the cross-phaseogram methods of Skoe et al. (2011), we assessed the difference in phase between responses in quiet and those in noise. Because the response in quiet was expected to occur earlier than the response in noise (Anderson et al., 2010a; Song et al., 2011a), the response in noise was predicted to phase lag the response in quiet. Inspection of the difference in phase between the responses over frequency and across time and an *a priori* hypothesis that differences might be largest for lower harmonics (Anderson et al.,

2010b; Song et al., 2011c) revealed that the largest phase differences between quiet and noise in Year 1 were found in the formant transition region of the response (20–60 ms) in the low harmonics (0–200 Hz). The average phase difference between responses in quiet and those in noise was calculated over this region for Year 1 and Year 2.

2.3.2. Within-session consistency

The consistency of the response within a recording session was determined by calculating the straight correlation between the two 3000 sweep replications generated for each recording (/da/ in quiet and /da/ in noise) over the entire response (0–180 ms). The resulting correlation coefficients were Fisher transformed prior to statistical analyses. Correlations reported were converted to r-values after statistical analyses.

2.3.3. Across-year consistency

Similar to the within-session consistency measure, the across-year consistency measure was calculated as the straight correlation of an individual's response from Year 1 with their response from Year 2 for both /da/ in quiet and /da/ in noise over the entire response (0–180 ms). The resulting correlation coefficients were Fisher transformed prior to averaging across the group. Correlations reported were converted to r-values after statistical analyses.

2.3.4. Response amplitude
2.3.4.1. Signal-to-noise ratio.
Root mean square amplitude over the entire response (0–180 ms) served as a measure of overall response amplitude for both /da/ in quiet and /da/ in noise. Signal-to-noise ratios were calculated by dividing these overall response amplitudes by the amplitudes calculated over the pre-stimulus period (−40 to 0 ms) for each response. We elected not to include an analysis of raw response amplitude because it is known to be unreliable in clinical evaluations of subcortical function and may be influenced by a number of subject factors that are not related to maturation of the nervous system, such as head size (Ponton et al., 2000; Salamy, 1984).

2.3.4.2. Spectral amplitudes of responses.
Spectral amplitudes of responses to /da/ in quiet and in noise were calculated using Fast Fourier Transforms conducted over the formant transition region of the response (20–60 ms) and averaging over 40 Hz wide bins centered at the fundamental frequency (F_0) and its integer harmonics through 1000 Hz (i.e., harmonics 2–10).

2.4. Statistics

Multivariate Repeated Measures ANOVAs were used for the analysis of response latencies and spectral amplitudes, while Repeated Measures ANOVAs were used for the analysis of signal-to-noise ratio and within-session consistency. In all cases, time (Year 1, Year 2) and condition (quiet, noise) served as within-subjects factors and main effects and interactions are reported. Follow-up paired t-tests were used when appropriate and for the comparison of the phase shift between responses. Because we hypothesized little change in response measures from Year 1 to Year 2, a conservative approach was to let α equal 0.05, allowing the greatest possibility of finding a significant result (contrary to our hypothesis). Spearman's correlations were also conducted between Year 1 and Year 2 to assess the reliability of the response measures.

3. Results

Speech-evoked brainstem response measures were highly replicable over one year. No change was found for any measure

from Year 1 to Year 2 (see Table 1). As would be predicted, a main effect of condition was found for all analyses, with responses to /da/ in quiet occurring earlier and being larger than responses to /da/ in noise. Because test-retest effects were the focus of the manuscript, these main effects of condition will not be discussed.

3.1. Consistency of response timing

From Year 1 to Year 2, there was no difference in response latencies in quiet and noise ($F_{6,19} = 0.875$, $p > 0.50$) and no interaction between condition and time ($F_{6,19} = 0.705$, $p > 0.65$), indicating that there is no change in the quiet-to-noise timing shift of response peaks over the course of one year of growth (see Fig. 1A, B and E and Table 1).

The lack of an interaction between time and condition was further supported by a lack of change from Year 1 to Year 2 in average response phase lag between responses to /da/ in quiet and /da/ in noise during the formant transition across the low harmonics ($t_{24} = 0.475$, $p > 0.60$; see Table 1).

3.2. Consistency of responses within the recording session and across years

The within-session consistency of the responses did not change from Year 1 to Year 2. Correlation strength between responses from the first half of the recording and those from the second half in Year 2 was not different from the split-half correlation of responses in Year 1 for the responses to /da/ in quiet and noise ($F_{1,24} = 0.769$, $p > 0.389$; see Fig. 1F and Table 1). Additionally there was no interaction between time and condition ($F_{1,24} = 3.055$, $p > 0.05$).

The across-year consistency correlations were 0.799 (SD = 0.29) for responses to /da/ in quiet and 0.689 (SD = 0.28) for responses to /da/ in noise, suggesting that overall the responses from the two years were very similar.

3.3. Signal-to-noise ratio

There was no difference from Year 1 to Year 2 in the signal-to-noise ratio of the response for responses in quiet or in noise ($F_{1,24} = 0.426$, $p > 0.50$) or interaction between time and condition ($F_{1,24} = 2.417$, $p > 0.13$). Although electrophysiological amplitudes are known to decrease with age (Ponton et al., 2000; Salamy, 1984), there is no indication of any change in response robustness as assessed by the signal-to-noise ratio.

3.4. Stability of frequency representation

For responses to /da/ in quiet and in noise, there was no change in spectral magnitudes from Year 1 to Year 2 ($F_{10,15} = 1.175$, $p > 0.35$) and no significant interaction of condition and time ($F_{10,15} = 1.199$, $p > 0.35$; see Fig. 1C and D and Table 1).

3.5. Reliability of response metrics

In order to assess reliability of the measures discussed here, Spearman's correlations were conducted between Year 1 and Year 2 values (see Table 1). Reliability estimates were generally good, although not for every measure, and appear to be larger for metrics that were automated (signal-to-noise ratio, spectral amplitudes, and within-session response replicability).

4. Discussion

The timing and spectral amplitudes of speech-evoked auditory brainstem response measures are remarkably consistent over one

Table 1
Means and standard deviations for response measures for Year 1, Year 2, and the test-retest change (computed as Year 2 minus Year 1) and reliability of measures (Year 1 correlated with Year 2). Measures that showed a significant change from Year 1 to Year 2 or a significant reliability correlation coefficient are bolded ($p < 0.05$).

	Year 1	Year 2	Change	Reliability
	Mean (Standard Deviation)			Spearman's rho
Timing				
Response Latencies in Quiet (ms)				
Peak 9	9.165 (0.48)	9.196 (0.37)	0.031 (0.62)	0.123
Trough 10	10.235 (0.60)	10.279 (0.48)	0.044 (0.79)	0.139
Peak 42	42.844 (0.25)	43.054 (0.47)	0.210 (0.46)	**0.565**
Trough 43	44.402 (0.75)	44.285 (0.71)	−0.117 (0.72)	**0.456**
Peak 52	52.889 (0.22)	52.985 (0.29)	0.096 (0.30)	**0.473**
Trough 53	54.004 (0.30)	54.112 (0.32)	0.108 (0.29)	**0.484**
Response Latencies in Noise (ms)				
Peak 9	9.914 (0.65)	9.906 (0.64)	−0.002 (0.99)	−0.185
Trough 10	10.748 (0.61)	10.712 (0.73)	−0.024 (1.01)	−0.154
Peak 42	43.682 (0.76)	43.810 (0.83)	0.096 (0.73)	**0.566**
Trough 43	45.050 (0.86)	45.298 (0.84)	0.202 (0.86)	**0.401**
Peak 52	53.142 (0.47)	53.265 (0.58)	0.054 (0.54)	**0.590**
Trough 53	54.284 (0.66)	54.415 (0.82)	0.094 (0.92)	0.305
Quiet-to-Noise Phase Shift (π radians)				
Low Harmonics	0.531 (0.39)	0.481 (0.43)	−0.041 (0.43)	0.355
Within-Session Replicability (r)				
Quiet	0.724 (0.30)	0.707 (0.29)	−0.035 (0.23)	**0.664**
Noise	0.534 (0.30)	0.591 (0.29)	0.086 (0.24)	**0.667**
Amplitude (signal-to-noise ratio)				
Quiet SNR	2.54 (0.63)	2.37 (0.66)	−0.171 (0.65)	**0.752**
Noise SNR	1.92 (0.54)	1.96 (0.56)	0.058 (0.49)	**0.601**
Spectral Encoding (μV)				
<u>Quiet</u>				
F0	0.1126 (0.036)	0.1030 (0.041)	−0.0112 (0.024)	**0.815**
H2	0.0614 (0.023)	0.0565 (0.022)	−0.0048 (0.017)	**0.662**
H3	0.0275 (0.113)	0.0252 (0.012)	−0.0024 (0.014)	0.319
H4	0.0192 (0.010)	0.0154 (0.007)	−0.0036 (0.011)	0.339
H5	0.0213 (0.010)	0.0193 (0.009)	−0.0024 (0.009)	**0.586**
H6	0.0126 (0.007)	0.0116 (0.006)	−0.0012 (0.007)	**0.510**
H7	0.0089 (0.006)	0.0083 (0.005)	−0.0012 (0.007)	**0.740**
H8	0.0056 (0.003)	0.0050 (0.002)	0.0004 (0.008)	0.202
H9	0.0058 (0.003)	0.0060 (0.004)	0.0004 (0.005)	**0.540**
H10	0.005 (0.005)	0.0052 (0.004)	1.39e-19 (0.006)	0.358
<u>Noise</u>				
F0	0.0617 (0.024)	0.0689 (0.027)	0.0029 (0.022)	**0.656**
H2	0.0374 (0.022)	0.0392 (0.020)	0.0013 (0.028)	0.231
H3	0.0189 (0.010)	0.0165 (0.008)	−0.0025 (0.015)	−0.195
H4	0.0163 (0.009)	0.0142 (0.006)	−0.0013 (0.014)	−0.117
H5	0.0129 (0.008)	0.0107 (0.006)	−0.0013 (0.007)	0.328
H6	0.0081 (0.005)	0.0076 (0.003)	0.0004 (0.005)	**0.429**
H7	0.0052 (0.003)	0.0052 (0.003)	−0.0008 (0.006)	0.336
H8	0.0038 (0.002)	0.0041 (0.002)	0.0013 (0.005)	0.142
H9	0.0053 (0.003)	0.0043 (0.002)	−0.0008 (0.005)	**0.511**
H10	0.0044 (0.004)	0.0039 (0.002)	−0.0008 (0.005)	**0.598**

year of growth in school-age children. The stability and consistency of speech-evoked auditory brainstem responses is clinically and theoretically important because response measures such as these have been linked to learning and communication skills in children and been shown to be malleable with short-term and life-long experience with sound (Anderson et al., 2010a; Banai et al., 2009; Hornickel et al., 2011; Krishnan et al., 2005; Russo et al., 2010; Song et al., 2011b; Wong et al., 2007). The results here suggest that speech-evoked ABRs are a possible tool for clinical and research assessment of auditory-based communication skills and experience-dependent neuroplasticity.

The clinically-viable reliability of the click-evoked ABR is well established (Edwards et al., 1982; Hood, 1998; Lauter and Oyler, 1992; Musiek et al., 2007; Sininger, 2007; Tusa et al., 1994). The highly replicable, consistent response pattern evoked by click stimuli and characteristic response changes with alterations of intensity or presentation rate underlies the use of the click-evoked brainstem response in infant hearing screening and assessments of central

auditory function (Hall, 2006; Hood, 1998; Musiek et al., 2007; Sininger, 2007). The reliability and replicability of ABRs to speech has been shown in adults (Song et al., 2011a) and here we confirm a similar consistency of responses in children over one year of growth.

Song et al. (2011a) assessed retest reliability over the course of three months in adults and reported that response measures such as overall amplitude, signal-to-noise ratio, response peak latencies, and spectral amplitudes were stable. The reliability coefficients of these measures were robust (Song et al., 2011c). Other work investigating training-related improvements in children suggested that responses from children are replicable and consistent, due to the lack of significant change in control groups from pre-test to post-test (Russo et al., 2005), similar to reliability shown in adults (Carcagno and Plack, 2011; Song et al., 2011a). The present study expanded those results to a larger set of response measures and found similar replicability in children as in adults. Importantly, unlike the previous training studies, the test-retest period in the current study was an entire calendar year on average, which
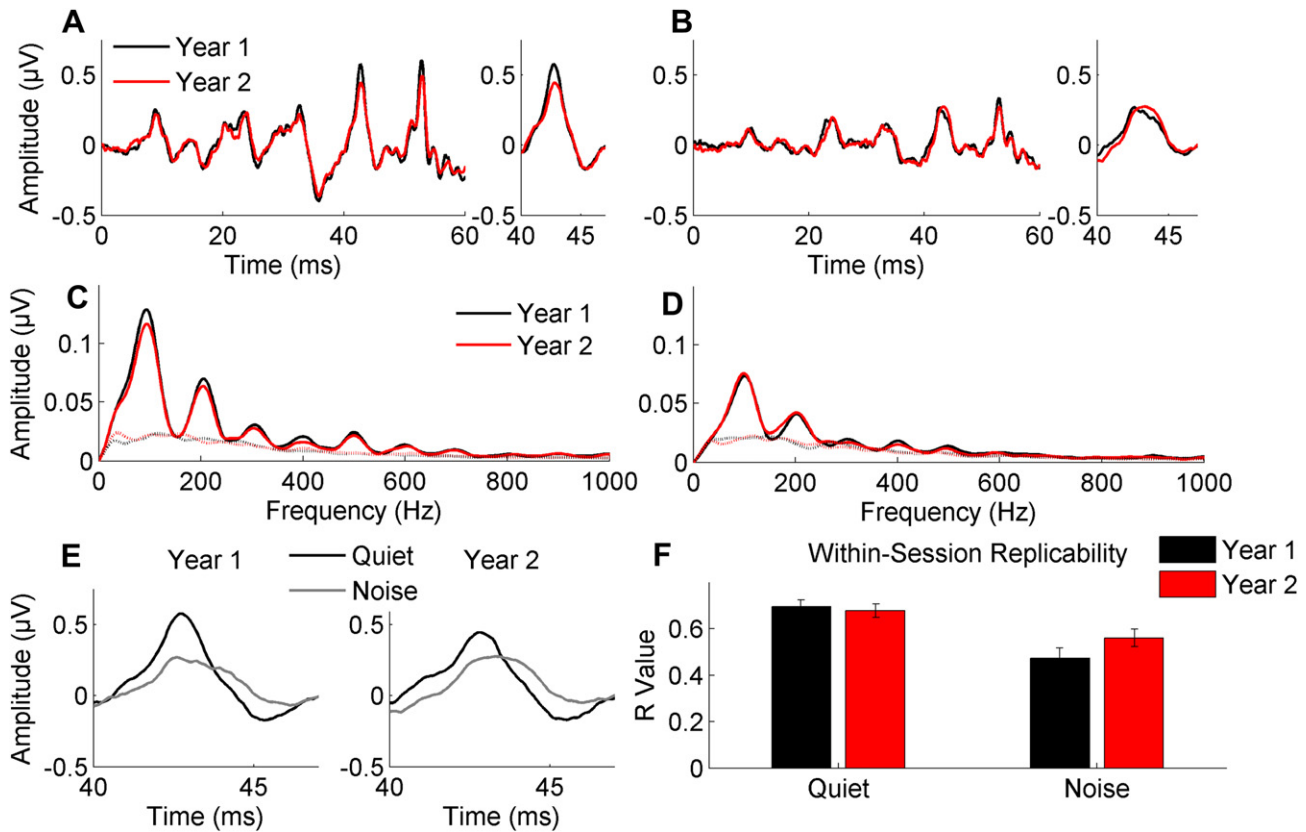
Fig. 1. Speech-evoked brainstem response timing and spectral encoding is stable over the course of one year in school-age children. Response peak latencies during the formant transition region (0–60 ms) do not change over the course of one year for responses to /da/ presented in quiet (A) or /da/ presented in noise (B). See insets for greater detail. The magnitude of the timing shift between responses in quiet and responses in noise, with responses in noise occurring later, is also stable (E). Spectral amplitudes of responses to /da/ in quiet (C) and /da/ in noise (D) also do not change over the course of one year. Additionally the consistency of the responses within the recording session is stable over one year of growth (F). The predicted main effect of condition was replicated for all measures, with responses in quiet being larger, earlier, and more consistent than responses in noise (A–F). Note: Year 1 is marked by black and Year 2 by red online and gray in print for all plots except (E) for which responses in quiet are marked by black and responses in noise by gray. Noise floors for plots C and D are marked by hashed lines.

increased the chances of detecting maturational or experiential change if it occurred. In the current study no changes in response timing measures or in spectral encoding were found over the course of one year. The consistency of response timing and spectral measures over one year of growth in the school-age years suggests that speech-evoked brainstem responses are at a maturational plateau in this age group, a theory that is supported by developmental studies.

The click-evoked ABR matures quickly, reaching adult-like response morphology at approximately two years of age (Hood, 1998; Salamy, 1984; Sininger, 2007). That the click-evoked response is measurable in neonates suggests that the basic neural structure of the auditory pathway is established by birth, congruent with the onset of hearing at approximately 20 weeks in utero (Graven and Browne, 2008), and supports the use of click-evoked responses to detect peripheral auditory function and nervous system pathologies (Hall, 2006; Hood, 1998; Sininger, 2007). The speech-evoked ABR, on the other hand, does not appear to mature until at least age five. Speech-evoked ABRs of five-year-old children are not significantly different than responses from school-age children (8–12 years old) while responses of children who are three and four years old differ in response morphology and timing (Johnson et al., 2008). The slower maturation of speech-evoked responses may be due to the acoustic complexity of speech relative to click stimuli, requiring more continuous and synchronous phasic activity from a variety of neural nuclei. Within-subject variability in click-evoked responses is greatest for young

children (5–7 years old), reduced for older children (10–12 years old), and least for adults (Lauter and Oyler, 1992), suggesting that neural synchrony and consistency is increasing during the same period that speech-evoked responses are maturing.

In addition to demonstrating that speech-evoked ABR measures do not significantly differ over the course of one year we estimated the reliability of these measures. In general reliability estimates were adequate, although reliability was sometimes low for individual response peaks or spectral amplitudes. Reliability coefficients above 0.7–0.8 are typically preferred (Nunnally, 1959) and were found for speech-evoked auditory brainstem response measures collected in adults (Song et al., 2011). As stated above, there is evidence that the variability of auditory brainstem responses decreases with age (Lauter and Oyler, 1992) and the weaker reliability estimates observed here relative to previous estimates in adults may be due to slightly greater response variability in this group of school-age subjects. Additionally, the test-retest period is one full year in the present study which is substantially longer than test-retest periods of months or days in previous studies showing speech-evoked response reliability (Russo et al., 2004; Song et al., 2011), click-evoked response reliability (Edwards et al., 1982; Lauter and Oyler, 1992), and test-retest reliability for well-utilized behavioral assessments of learning and achievement (McGrew and Woodcock, 2001; Torgesen et al., 1999; Wagner et al., 1999). Click-evoked response reliability is weaker for responses collected two years apart than responses collected on the same day (Tusa et al., 1994) and it is very likely that the reliability

estimates in the present study would be higher if the test-retest interval was shorter.

Altering stimulus and recording parameters, such as stimulus presentation rate and response sampling rate, may also lead to enhancements of reliability. While the fast sampling rate used here allows for finer temporal resolution, it also reveals the effects of very subtle differences in external factors, such as slight differences in electrode impedance. Additionally, the extremely high temporal resolution may impact the subjective peak picking analyses by reflecting these subtle differences in external factors that can impact response morphology. If speech-evoked ABRs are to be used for clinical diagnoses, manipulations of the stimulus and recording paradigms should be made to maximize reliability and objective response metrics should be relied upon as the key measures.

Despite the long test-retest interval and potentially greater variability in responses due to age and external factors, no changes were seen in response timing and spectral encoding over the course of one year of growth, highlighting that within-subject variability is minimal and responses are consistent from test to retest. Additionally, the straight correlations between responses from Year 1 and Year 2 were on average 0.7 and higher for both /da/ collected in quiet and /da/ collected in noise, indicating that the responses themselves were highly consistent across years and lower reliabilities seen for some of the reported measures, such as peak latencies, may be due to the more subjective nature of the analyses. It is important to note that automated response measures, particularly spectral magnitudes and within-session consistency, appear to be more reliable than measures requiring manual identification of response peaks, suggesting they may be preferred in a clinical setting.

The lack of test-retest changes found here strengthens the notion that changes in neural function following auditory training in school-age children reflect fundamental changes in sensory processing associated with meaningful engagement with sound, and not test-retest factors. Although auditory brainstem responses are known to be malleable with meaningful or impaired interaction with sound (Anderson et al., 2010a,b; Banai et al., 2009; Hornickel et al., 2009, 2011; Kraus and Chandrasekaran, 2010; Krishnan et al., 2005; Musacchia et al., 2007; Wible et al., 2004; Wong et al., 2007; Xu et al., 2006), in the absence of such active engagement they are expected to be stable. If positive changes are found in the speech-evoked response measures included here, which have been linked to auditory-based communication skills (Anderson et al., 2010a; Banai et al., 2009; Hornickel et al., 2011), it is possible that behavioral improvements may follow. Additionally, these results suggest that differences in speech-evoked response measures between groups of school-age children, for example good and poor readers, are not due to a difference in maturational differences or response inconsistencies. The observed consistency of the speech-evoked ABR during the school-age years and repeat evidence of relationships between response measures and behavioral measures of communication skills suggests that speech-evoked brainstem responses could possibly serve as a unique metric of biological correlates of auditory-based communication skills and experience-dependent modulation of auditory function.

5. Conclusion

The speech-evoked brainstem response is replicable and consistent over one year of growth in typically-developing children. These results mirror recent analyses of test-retest reliability of speech-evoked brainstem response measures in adults showing that measures of response timing and frequency representation are reliable across months. Our expansion of these results to typically-developing children assessed over one year supports a theory that speech-evoked brainstem responses are at a maturational plateau

in this age group and may contribute unique information for research and clinical assessment of auditory processing within groups of school-age children. Future studies should investigate the consistency of these response measures in children with auditory-based communication and/or learning deficits to determine if children with impairments show a similar response consistency as seen in typically-developing children or whether response variability is a biological hallmark of clinical impairments.

Acknowledgments

References

Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245, 35–47.

Anderson, S., Skoe, E., Chandrasekaran, B., Kraus, N., 2010a. Neural timing is linked to speech perception in noise. J. Neurosci. 30, 4922–4926.

Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S.G., Kraus, N., 2010b. Brainstem correlates of speech-in-noise perception in children. Hear. Res. 270, 151–157.

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S.G., Kraus, N., 2009. Reading and subcortical auditory function. Cereb. Cortex 19, 2699–2707.

Campbell, T., Kerlin, J.R., Bishop, C.W., Miller, L.M., 2011. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear Hear. doi:10.1097/AUD.0b013e3182280353.

Carcagno, S., Plack, C.J., 2011. Subcortical plasticity following perceptual learning in a pitch discrimination task. J. Assoc. Res. Otolaryngol. 12, 89–100.

Edwards, R.M., Buchwald, J.S., Tanguay, P.E., Schwafel, J.A., 1982. Sources of variability in auditory brain stem evoked potential measures over time. Electroencephalogr. Clin. Neurophysiol. 53, 125–132.

Gorga, M., Abbas, P., Worthington, D., 1985. Stimulus calibration in ABR measurements. In: Jacobsen, J. (Ed.), The Auditory Brainstem Response. College-Hill Press, San Diego, pp. 49–62.

Graven, S.N., Browne, J.V., 2008. Auditory development in the fetus and infant. Newborn Infant Nurs. Rev. 8, 187–193.

Hall, J.W., 2006. New Handbook of Auditory Evoked Responses. Allyn & Bacon, Boston, MA.

Hood, L.J., 1998. Clinical Applications of the Auditory Brainstem Response. Delmar Learning, Clifton Park, NY.

Hornickel, J., Chandrasekaran, B., Zecker, S.G., Kraus, N., 2011. Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. Behav. Brain Res. 216, 597–605.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S.G., Kraus, N., 2009. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. Proc. Natl. Acad. Sci. U.S.A. 106, 13022–13027.

Issa, A., Ross, H.F., 1995. An improved procedure for assessing ABR latency in young subjects based on a new normative data set. Int. J. Pediatr. Otorhinolaryngol. 32, 35–47.

Johnson, K., Nicol, T., Kraus, N., 2008. Developmental plasticity in the human auditory brainstem. J. Neurosci. 28, 4000–4007.

Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67, 971–995.

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

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

Lauter, J.L., Oyler, R.F., 1992. Latency stability of auditory brainstem responses in children aged 10–12 compared with younger children and adults. Br. J. Audiol. 26, 245–253.

McGrew, K.S., Woodcock, R.W., 2001. Technical Manual: Woodcock-Johnson III. Riverside Publishing, Itasca, IL.

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

Musiek, F., Shinn, J.B., Jirsa, R.E., 2007. The auditory brainstem response in auditory nerve and brainstem dysfunction. In: Burkard, R.F., Eggermont, J.J., Don, M. (Eds.), Auditory Evoked Potentials: Basic Principles and Clinical Application. Lippincott Williams & Wilkins, Baltimore, MD, pp. 291—312.

Nunnally, J.C., 1959. Tests and measurements: Assessment and prediction. McGraw Hill, New York.

Ponton, C.W., Eggermont, J.J., Kwong, B., Don, M., 2000. Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. Clin. Neurophysiol. 111, 220—236.

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

Russo, N., Nicol, T., Zecker, S.G., Hayes, E., Kraus, N., 2005. Auditory training improves neural timing in the human brainstem. Behav. Brain Res. 156, 95—103.

Russo, N., Hornickel, J., Nicol, T., Zecker, S.G., Kraus, N., 2010. Biological changes in auditory function following training in children with autism spectrum disorders. Behav. Brain Funct. 6, 60.

Salamy, A., 1984. Maturation of the auditory brainstem response from birth through early childhood. J. Clin. Neurophysiol. 1, 293—329.

Sininger, Y.S., 2007. The use of auditory brainstem response in screening for hearing loss and audiometric threshold prediction. In: Burkard, R.F., Eggermont, J.J., Don, M. (Eds.), Auditory Evoked Potentials: Basic Principles and Clinical Application. Lippincott Williams & Wilkins, Baltimore, MD, pp. 254—274.

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

Skoe, E., Nicol, T., Kraus, N., 2011. Cross-phaseogram: objective neural index of speech sound differentiation. J. Neurosci. Methods 196, 308—317.

Song, J.H., Nicol, T., Kraus, N., 2011a. Test-retest reliability of the speech-evoked auditory brainstem response. Clin. Neurophysiol. 122, 346—355.

Song, J.H., Skoe, E., Banai, K., Kraus, N., 2011b. Training to improve hearing speech in noise: biological mechanisms. Cerebral Cortex 122, 1890—1898.

Song, J.H., Skoe, E., Banai, K., Kraus, N., 2011c. Perception of speech in noise: neural correlates. J. Cogn. Neurosci. 23, 2268—2279.

Syntrillium Software: Syntrillium Software Corporation, 2003. Cool Edit Pro, the Complete Multitrack Digital Audio Recorder, Editor, Mixer, Version 2.1, San Jose, CA.

Torgensen, J.K., Wagner, R.K., Rashotte, C.A., 1999. Test of Word Reading Efficiency (TOWRE). Pro-Ed, Austin, TX.

Torgesen, J.K., Wagner, R.K., Rashotte, C.A., 1999. Examiner's Manual: Test of Word Reading Efficiency. Pro-Ed, Austin, TX.

Tusa, R.J., Stewart, W.F., Shechter, A.L., Simon, D., Liberman, J.N., 1994. Longitudinal stud of brainstem auditory evoked responses in 87 normal human subjects. Neurology 44, 528—532.

Wagner, R.K., Torgesen, J.K., Rashotte, C.A., 1999. Examiner's Manual: The Comprehensive Test of Phonological Processing. Pro-Ed, Austin, TX.

Wible, B., Nicol, T., Kraus, N., 2004. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biol. Psychol. 67, 299—317.

Woerner, C., Overstreet, K., 1999. Wechsler Abbreviated Scale of Intelligence (WASI). The Psychological Corporation, San Antonio, TX.

Wong, P.C.M., Skoe, E., Russo, N., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420—422.

Xu, Y., Krishnan, A., Gandour, J., 2006. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17, 1601—1605.