Hearing Research 287 (2012) 3–5

Contents lists available at SciVerse ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Letter to the Editor

Reliability of the auditory brainstem responses to speech over one year in school-age children: A reply to Drs. McFarland and Cacace

In Hornickel, Knowles, and Kraus (Hornickel et al., 2012), we provide evidence that the speech-evoked auditory brainstem response (speech-ABR) in typically-developing eight to thirteen year-old children is stable over the course of one year. Our comprehensive study of the reliability of multiple speech-ABR measures yielded reliabilities ranging from 0.11 to 0.82 (reprinted in Table 1 column 1). While reliabilities higher than 0.7 are preferred for clinical diagnosis (Nunnally, 1959), suitable reliability for group analyses is 0.6 (Salvia and Ysseldyke, 2004). Drs. McFarland and Cacace have raised concerns that the reliabilities reported in Hornickel et al. (2012) were low which, given the large number of measures reported, increased the likelihood of false alarms if used diagnostically. While we understand their concerns, our intent was to be inclusive in our reporting of reliability and we acknowledge that when the situation demands it, such as in clinical use, one would choose a small number of measures with the highest reliability.

The reliability of the speech-ABR will probably never reach the levels seen for click-evoked brainstem responses for a variety of reasons, including the comparatively richer, spectrotemporally-dynamic information present in speech compared to clicks and the experience-dependent plasticity of the speech-ABR. However, it is also likely that both the age of the subjects and the retest interval contributed to test–retest variance in Hornickel et al. (2012). The data were collected over a one year test–retest interval, in contrast to the two week interval typical for diagnostic behavioral assessments (McGrew and Woodcock, 2001; Salvia and Ysseldyke, 2004; Torgensen et al., 1999; Wagner et al., 1999). As may be expected, reliability of behavioral measures decreases as the retest interval increases (McGrew and Woodcock, 2001). Previous studies of speech-ABRs used test–retest intervals only a few months in length (Russo et al., 2004; Song et al., 2011). Moreover, ABRs of children are known to be more variable than adults both between and within subjects (Lauter and Oyler, 1992), and click-evoked reliability decreases with longer test–retest intervals (Tusa et al., 1994). Thus, a longer test–retest interval, a younger subject population, and a complex stimulation all contribute to test–retest variability.

Due to our modest sample size ($n = 26$), we elected to assess reliability using Spearman's correlations. While we agree with Drs. McFarland and Cacace that rank ordering the data for reliability estimates might be more appropriate for rank ordered clinical measures, Spearman's correlations are the more conservative estimate. Reliabilities are largely the same when calculated as Pearson's correlations (compare Table 1, columns 1 and 2). In fact, a greater number of speech-ABR measures show significant correlations and reliabilities above 0.7 when using the Pearson's *r*. This suggests that Spearman's correlations in some cases underestimated the reliability of the measures.

Reliability estimates in Hornickel et al. (2012) may also suffer from a restriction of range (Salvia and Ysseldyke, 2004). As Drs. McFarland and Cacace suggest, norms collected in children with a wide range of reading ability would be most informative if the speech-ABR is to be used with clinical populations. We expand on our original publication by presenting additional data from 19 learning impaired children who did not differ in age from the participants in Hornickel et al. (2012) ($t_{43} = 1.047, p > 0.3$). Overall reliabilities are generally the same in this broader dataset (compare Table 1, columns 2 and 3), suggesting that these speech-ABR measures are consistent across a group of children with a wide range of reading ability.

While we maintain that the speech-evoked auditory brainstem response contributes valuable information to the assessment of auditory processing, we do not advocate use of the particular stimulus; presentation, recording, and processing parameters; or recording hardware and software utilized in Hornickel et al. (2012) for immediate clinical purposes. We expect that reliability can be strengthened with refinements to the stimulus and recording parameters. We are in the process of designing such a clinically-oriented protocol with a different stimulus, and it includes a small subset of the 37 response measures reported in Hornickel et al. (2012), namely the specific timing and frequency-domain measures previously shown to be discriminative of reading and/or speech-in-noise perception abilities (Anderson et al., 2010a, 2010b; Banai et al., 2009; Hornickel et al., 2009; Wible et al., 2004). When considering only this subset of measures (timing of Peak V, Trough A, and Trough O; spectral amplitudes), reliability coefficients are on the order of 0.7–0.9 (see Table 2) for both the original typically-developing cohort and the expanded dataset including learning impaired children. In addition to having higher reliabilities, our selection of a targeted subset of clinically-oriented measures also reduces the possibility of false alarms.

It is important to note that the weakest reliabilities in Hornickel et al. (2012) and the expanded dataset are seen for the manually-identified peak and trough latencies, which are most susceptible to the subjective judgments of raters. Objective indices, such as the within-session consistency and spectral encoding measures, have reliabilities approaching or exceeding 0.7 for responses in both quiet and noise. We also measured the stability of the speech-ABR waveform morphology by correlating responses from Year 1 and Year 2 for responses in quiet and noise (respectively) and found that the speech-ABRs across the two years share much

0378-5955/$ – see front matter
doi:10.1016/j.heares.2012.03.014

Table 1
Reliabilities for speech-evoked brainstem response measures to a 170 ms /da/ presented in quiet and noise from 26 typically-developing children from Hornickel et al. (2012) as assessed by Spearman's correlations (column 1) and Pearson's correlations (column 2). Reliabilities, as calculated by Pearson's correlations, are also provided for an expanded dataset that includes 19 learning impaired children of the same age (column 3). Significant correlations, $p < 0.05$, are bolded.

	Reliability		
	Spearman's rho ($n = 26$)	Pearson's r ($n = 26$)	With 19 learning impaired children, Pearson's r ($n = 26$)
Timing			
Response latencies in quiet			
Peak 9	0.123	−0.034	−0.012
Trough 10	0.139	−0.056	0.011
Peak 42	**0.565**	0.328	0.218
Trough 43	**0.456**	**0.510**	**0.449**
Peak 52	**0.473**	0.343	**0.528**
Trough 53	**0.484**	**0.563**	0.271
Response latencies in noise			
Peak 9	−0.185	−0.150	−0.146
Trough 10	−0.154	−0.096	−0.174
Peak 42	**0.566**	**0.581**	**0.340**
Trough 43	**0.401**	**0.481**	0.235
Peak 52	**0.590**	0.360	**0.371**
Trough 53	0.305	0.250	0.153
Quiet-to-noise phase shift			
Low Harmonics	0.355	**0.449**	0.022
Within-session replicability			
Quiet	**0.664**	**0.709**	**0.741**
Noise	**0.667**	**0.699**	**0.673**
Amplitude (signal-to-noise ratio)			
Quiet SNR	**0.752**	**0.500**	**0.591**
Noise SNR	**0.601**	**0.601**	**0.424**
Spectral encoding			
Quiet			
F0	**0.815**	**0.807**	**0.822**
H2	**0.662**	**0.682**	**0.542**
H3	0.319	0.360	0.035
H4	0.339	**0.464**	**0.416**
H5	**0.586**	**0.633**	**0.662**
H6	**0.510**	**0.582**	**0.626**
H7	**0.740**	**0.784**	**0.649**
H8	0.202	0.025	0.102
H9	**0.540**	**0.660**	**0.459**
H10	0.358	**0.800**	**0.686**
Noise			
F0	**0.656**	**0.632**	**0.622**
H2	0.231	0.175	0.074
H3	−0.195	−0.101	−0.199
H4	−0.117	−0.068	0.017
H5	0.328	**0.700**	**0.351**
H6	**0.429**	**0.582**	0.265
H7	0.336	0.362	**0.363**
H8	0.142	0.241	−0.088
H9	**0.511**	**0.530**	**0.341**
H10	**0.598**	**0.649**	**0.427**

of their variance (quiet: $r = 0.799$; noise: $r = 0.689$). Thus, the response waveforms are quite reliable from one year to the next and the weaker reliability for discrete peak and trough latencies is likely due to the more subjective nature of the analysis. We continue to develop new, objective analysis techniques that can capture the precision of auditory timing without the limitations of traditional methods like manual peak picking (Kraus, 2011; Skoe et al., 2011). Both this study and our continuing research indicate that automated, objective measures yield the highest reliability and might best be applied in the clinic.

Table 2
Reliabilities (Pearson's r) for responses to a 40 ms /da/ stimulus presented at 10.9 Hz for the same subjects in Hornickel et al. (2012) (column 1) and for the expanded dataset that includes 16 learning impaired children of the same age (column 2). Data from 2 participants in the original dataset and from 3 of the learning impaired children included in Table 1 were lost due to computer errors. Significant correlations, $p < 0.05$, are bolded.

	Reliability (Pearson's r)	
	From Hornickel et al. (2012) ($n = 24$)	With 16 learning impaired children ($n = 40$)
Timing		
Response latencies in quiet		
Peak V	**0.569**	**0.779**
Trough A	**0.801**	**0.797**
Trough O	**0.516**	**0.437**
Neural synchrony		
VA slope	0.326	**0.503**
Spectral encoding		
Quiet		
F0 (103–120 Hz)	**0.821**	**0.789**
F1 (410–755 Hz)	**0.925**	**0.918**
HF (755–1130 Hz)	**0.855**	**0.774**

We want to conclude by defending our position that the speech-ABR be considered *as part* of an assessment battery of auditory processing. We have not, nor will ever, argue that the speech-evoked auditory brainstem response be used for dichotomous diagnostic decisions of reading or auditory processing impairments. Common practice is to make a diagnostic decision based on performance across a number of observations and measurements and, in that spirit, we maintain that the speech-ABR can provide a biological complement to a comprehensive assessment. Speech-ABR measures not only reflect cognitive communication skills such as reading and hearing in noise (Anderson et al., 2010a, 2010b; Banai et al., 2009; Hornickel et al., 2009; Wible et al., 2004), but reliability indices reported here are comparable to behavioral tests commonly used in auditory processing batteries, such as the Dichotic Digits, Pitch Patterns, Selective Auditory Attention, Competing Sentences, and SCAN tests, which range from approximately 0.55 to 0.99, and are generally near 0.7 (Amos and Humes, 1998; Bakker et al., 1978; Domitz and Schow, 2000; Keith, 2000).

We thank Drs. McFarland and Cacace for their feedback and their continuing efforts to validate the test–retest reliability of auditory processing assessments, particularly for clinically-oriented measures (McFarland and Cacace, 2002; McFarland and Cacace, 2003; McFarland and Cacace, 2011; McFarland and Cacace, 2012). We maintain that the auditory brainstem response to speech is an objective measure of auditory processing that captures rich, multi-dimensional information about the biological processing of complex sounds. Although clinically-oriented techniques are still being refined, analyses of the utility and consistency of the speech-evoked ABR within individuals is encouraging.

Acknowledgments

We would like to thank Erika Skoe and Trent Nicol for their review of the manuscript.

References

Amos, N.E., Humes, L.E., 1998. SCAN test–retest reliability for first- and third-grade children. J. Speech Lang. Hear. Res. 41, 834–845.

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.

Bakker, D.J., Van der Vlugt, H., Claushuis, M., 1978. The reliability of dichotic ear asymmetry in normal children. Neuropsychologia 16, 753–757.

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

Domitz, D.M., Schow, R.L., 2000. A new CAPD battery-multiple auditory processing assessment: factor analysis and comparisons with SCAN. American J. Audiol. 9, 101–111.

Hornickel, J., Knowles, E., Kraus, N., 2012. Test–retest consistency of speech-evoked auditory brainstem responses in typically-developing children. Hear. Res. 284, 52–58.

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. 106, 13022–13027.

Keith, R.W., 2000. Development and standardization of SCAN-C test for auditory processing Disorders in children. J. Am. Acad. Audiol. 11, 438–445.

Kraus, N., 2011. Listening in on the listening brain. Phys. Today 64, 40–45.

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.

McFarland, D., Cacace, A.T., 2002. Factor analysis in CAPD and the "unimodal" test battery: do we have a model that will satisfy? Am. J. Audiol. 11, 7–9.

McFarland, D., Cacace, A.T., 2003. Potential problems in the differential diagnosis of (central) auditory processing disorder (CAPD or APD) and attention-deficit hyperactivity disorder (ADHD). J. Am. Acad. Audiol. 14, 278–280.

McFarland, D., Cacace, A.T., 2011. Covariance is the proper measure of test-retest reliability. Clin. Neurophysiol. 122, 1893.

McFarland, D., Cacace, A.T., 2012. Questionable reliability of the speech-evoked auditory brainstem response (sABR) in typically-developing children. Hear. Res. 287, 1–2.

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

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

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

Salvia, J., Ysseldyke, J.E., 2004. Assessment in Special and Inclusive Education, nineth ed.. Houghton Mifflin, Boston, MA.

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., 2011. Test-retest reliability of the speech-evoked auditory brainstem response. Clin. Neurophysiol. 122, 346–355.

Torgensen, 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.

Jane Hornickel*, Erica Knowles
Auditory Neuroscience Laboratory, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA

Nina Kraus
Auditory Neuroscience Laboratory, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA

Department of Neurobiology and Physiology, Northwestern University, 633 Clark Street, Evanston, IL 60208, USA

Department of Otolaryngology, Northwestern University, 633 Clark Street, Evanston, IL 60208, USA

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

21 March 2012