https://doi.org/10.1044/2020_JSLHR-20-00383

Journal of Speech, Language, and Hearing Research

Comparisons of auditory brainstem responses between a laboratory and simulated home environment

Ashley Parker[1,2,3], Candace Slack[1,3], Erika Skoe[1,2,3†]

Affiliations:

[1]Department of Speech, Language, and Hearing Sciences; [2]Connecticut Institute for Brain and Cognitive Sciences; [3]University of Connecticut, Storrs, CT 06269 USA

†Corresponding author:

Erika Skoe, Ph.D.

University of Connecticut

2 Alethia Drive, U-1085, Storrs, CT 06269

860-486-3685

Email: erika.skoe@uconn.edu

ABSTRACT

PURPOSE

Miniaturization of digital technologies has created new opportunities for remote health care and neuroscientific fieldwork. The current study assesses comparisons between in-home auditory brainstem response (ABR) recordings and recordings obtained in a traditional lab setting.

METHOD

Click-evoked and speech-ABRs were recorded in 12 normal-hearing, young adult participants over three test sessions in: (1) a shielded sound booth within a research lab, (2) a simulated home environment, and (3) the research lab once more. The same single-family house was used for all home testing.

RESULTS

Analyses of ABR latencies, a common clinical metric, showed high repeatability between the home and lab environments across both the click-evoked and speech-ABRs. Like ABR latencies, response consistency (RC) and signal-to-noise ratio (SNR) were robust both in the lab and in the home, and did not show significant differences between locations, although variability between the home and lab was higher than latencies, with two participants influencing this lower repeatability between locations. RC and SNR also patterned together, with a trend for higher SNRs to pair with more consistent responses in both the home and lab environments.

CONCLUSION

Our findings demonstrate the feasibility of obtaining high-quality ABR recordings within a simulated home environment that closely approximate those recorded in a more traditional

recording environment. This line of work may open doors to greater accessibility to underserved clinical and research populations.

BACKGROUND

Electrobiological techniques are commonly used across fields of healthcare and research, with the capability of assessing function throughout the body, including the heart (electrocardiography), muscle contractions (electromyography), the stomach (electrogastrography), eyes (electrophotography), and the brain (electroencephalography—EEG). Traditionally, these measurements have been limited to clinical or laboratory settings, though the advent of small and portable equipment is creating the opportunity for in-home healthcare and research fieldwork. The goal of the current study is to compare EEG recordings—specifically, the auditory brainstem response (ABR)—between recordings collected in a simulated home setting and recordings collected from a traditional research laboratory setting, using standard clinical equipment that has small enough footprint to be portable and easily transported outside the lab.

Under the umbrella of EEG are auditory evoked potentials (AEPs), objective electrophysiological responses to sound representing activity along the auditory pathway. The ABR is a class of AEPs generated primarily from the auditory nerve and subcortical sources. The ABR, a low-voltage signal, is usually recorded by placing three to four electrodes on the scalp to record electrical activity to repeated stimulation from these deep sources, thus making it a far-field response (see Hall, 2007 for a full review). In comparison to other AEPs, ABRs have more mainstream clinical use. ABRs are routinely used by audiologists and other hearing health care professionals for newborn hearing screenings (e.g. Johnson et al., 2005), to estimate hearing

threshold sensitivity (e.g. Sininger et al., 1997), in determining hearing loss and type (e.g. Norton et al., 2000), in lesion detection (e.g. Achor & Starr, 1980), and for intraoperative monitoring (e.g. Daspit et al., 1982). The research community has given much attention to using the ABR to study the auditory correlates of language and reading ability and disability (e.g. Banai et al., 2009; Hornickel & Kraus, 2013; Neef, et al., 2017; Skoe et al., 2017), to examine the functional integrity of the early auditory neural pathway throughout the lifespan (e.g. Skoe et al., 2015), and more recently to study cochlear synaptopothy (e.g. Mehraei et al., 2016; Grose et al., 2017).

The most common stimulus for an ABR is a short click, known for its utility in estimating cochlear function in the 2-4 kHz region (Gorga et al., 1985; van der Drift et al., 1986). In humans, waves I, III, and V are the most prominent ABR waves, and reflect synchronized activity of the auditory nerve (wave I), cochlear nucleus (wave III), lateral lemniscus (wave V), and inferior colliculus (wave V) (Hall, 2007; Hood, 1998). Additionally, ABRs can be recorded to complex sounds, such as more naturalistic speech syllables (referred to here as the "speech-ABR", though it has received other labels in the literature as well—see Coffey et al., 2019 for a summary of common terminology and acronyms). The speech-ABR, which is considered to be a better approximation of speech processing than recordings to click stimuli (Skoe & Kraus, 2010), reflects predominately subcortical auditory system activity when recorded to frequencies in the range of the human voice (Bidelman, 2018).

The speech-ABR includes a transient onset response and sustained, phase-locked frequency-following response (FFR) to periodic acoustic elements (e.g., vowels). For the current investigation, we used a 40 ms speech stimulus /da/ to evoke speech-ABRs, with the advantage that the characteristics of both the stimulus and response are well-described in the literature (e.g. Johnson et al., 2008; Banai et al., 2009; Vander Werff & Burns, 2011; Jafari & Malayeri, 2014;

Malayeri et al., 2014; Skoe et al., 2015). This stimulus produces a series of highly-repeatable waves labeled V, A, C, D, E, F, and O. Waves V, A, and O (and wave C—not analyzed in this current study because of its poor reliability in the literature) are considered transient responses corresponding to transient stimulus features—the beginning (V, A) and end of voicing (O), respectively—while waves D, E, and F comprise the FFR and correspond to sustained features of the stimulus, namely the fundamental frequency (F_0) and harmonics within the consonant-vowel formant transition (reviewed in Skoe & Kraus, 2010). Dimensions of the speech-evoked ABR that are routinely analyzed include response latencies and amplitudes, response consistency (RC), signal-to-noise ratio (SNR), and spectral amplitudes of responses. While currently exclusive to research settings, speech-ABRs have been on the scene since the 1980s (Greenberg, 1980; Galbraith et al., 1995; Galbraith et al., 2004), with a movement to adopt them into clinical practice (Kraus et al., 2017; Ribas-Prats et al., 2018).

A clinical test such as the ABR, whether conducted in the home or within a laboratory or clinical setting, serves little use if the results are not stable and repeatable (i.e. high test-retest reliability). If the results of a test change significantly from session to session (or even block to block) without a biological change, their validity may not be dependable, especially as a diagnostic tool. As an extensively-used clinical tool, the test-retest reliability of the ABR has been heavily-studied in the literature for decades. In an early investigation of the test-retest reliability of the click-evoked ABR over an interval of several months, Edwards et al. (1982) found no significant differences in response amplitudes or latencies. Similarly, Oyler et al. (1991) found minimal variation in click-evoked ABR latencies between sessions conducted biweekly under three different stimulus conditions (monaural right, monaural left, and binaural).

Tusa et al. (1984) showed that within-session ABRs are more reproducible than are between-session (spaced two years apart), though neither show differences that are clinically significant.

ABR latencies, a measure of clinical utility, show high levels of reliability at retest, where deviations of only fractions of a millisecond are considered clinically significant (Hall, 2007). However, not all characteristics of the ABR are equally reliable. Some studies have shown that ABR amplitudes are more variable than latencies between tests (Dzulkarnain et al., 2014), even in normal-hearing ears (Schwartz et al., 1994). However, recent work in the study of noise-induced cochlear synaptopathy has brought attention specifically to the amplitude of wave I, which has shown high reliability within listeners when test sessions are spaced <1 week apart on average (Prendergast et al., 2018).

Speech-ABRs are also known for their reliability. Song et al. (2011) assessed reliability of amplitudes and latencies in speech-ABRs in quiet and in background noise, finding no significant differences at retest, and Bidelman et al. (2018) also found high reliability between four sessions in a one-month period. Moreover, Hornickel et al. (2012a) showed high repeatability of the speech-ABR's latency, RC, and SNR in school-aged children at two time points separated by one year. Still, there have been debates regarding the best methods for evaluating reliability in the ABR and broader EEG literature. Opinions on best approaches are varied, where not all are considered to be as equally valid at gauging reliability (McFarland & Cacace, 2011; McFarland & Cacace, 2012). This motivates our approach of using multiple methods to compare recordings between and within locations.

Traditionally, electrobiological measures, including ABRs, have been limited to laboratory and clinical settings due to the need for a sound booth and the large size of the equipment. However, EEG equipment is now becoming smaller and more mobile and the

hardware and software is less prone to electrical interference, allowing for more portability and remote testing in naturalistic environments (e.g. Debener et al., 2015; Zink et al., 2016; Krigolson et al., 2017; Kuziek et al., 2017; Kuziek et al., 2018; Lau-Zhu et al., 2019; Scanlon et al., 2019). Specific to the ABR, recent work has moved outside the lab by conducting recordings in the home (Tecoulesco et al., 2020), at sports-medicine clinics (Kraus et al., 2016) and in schools (Kraus et al., 2014; Krizman et al., 2015a).

Proof of reliability for in-home ABR recordings could expand options for audiology and offer accommodations for individuals who may have a difficult time getting to a clinic. Some individuals may prefer in-home healthcare to commuting to a lab, clinic, or hospital for a variety of reasons, including, but not limited to, a lack of transportation options, restricted schedules, cost, motor dysfunction or disabilities (e.g., Parkinson's disease, arthritis), poor vision (e.g., cataract), cognitive or communication disabilities (e.g., dementia, autism spectrum disorder, stroke), or equilibrium disorders (Bitterman, 2011). More generally, some patient groups such as young children, geriatrics, and other difficult-to-test populations may be more relaxed in the familiar environment of their home. Relaxation may lead to less muscle tension and movement artifact, allowing potentially more accurate results with also shorter test times. Many branches of healthcare already offer in-home or remote services ranging from nursing services (e.g. Giménez-Díez et al., 2020), physical (Calyam et al., 2016) and occupational (Bennett et al., 2019) therapy, speech-language pathology (Godlove et al., 2019), x-ray imaging (Toppenberg et al., 2019) and pharmaceutical services (Emblin et al., 2016).

The current study aims to compare ABR recordings in a home setting, where the setting is less controlled, to those collected in a research laboratory. This aim was motivated by recent work, including work from our own group, where recordings were made in settings such as

schools, locker-rooms, and home environments. In healthy young adults, we examine three components of the ABR: absolute latency, response consistency (RC), and signal-to-noise ratio (SNR). The experimental protocol involved a total of three test sessions: the first and last took place in a lab environment, and the middle session occurred in a single-family home, with all participants being tested in the same home. This design allowed for a comparison of within- and between-location repeatability.

Based on our research group's experience recording in non-traditional environments (e.g. Krizman et al., 2015b; Tecoulesco et al., 2020), we had grounds for predicting that the simulated home environment would yield high quality recordings with robust SNRs, and that compared to the lab environment, any differences in latency and RC between locations could be explained by slightly lower SNRs in the home environment from increased electrical noise and/or increased background noise.

METHOD

Participants

Twelve young adults (20-27 years, mean 24.84 years; 1 male, 11 females), all students at the University of Connecticut, participated in this study. Respondents to recruitment advertisements were screened via audiologic testing. All participants were confirmed to have clinically-normal hearing bilaterally (air conduction thresholds $\leq$ 20 dB HL for octave frequencies from 0.25 to 8 kHz; GSI 61 audiometer, Grason-Stadler, Inc.) (Figure 1) and passed an otoscopic exam and a distortion product otoacoustic emissions screening (Madsen Alpha OAE screener, Otometrics, Inc.). No participants needed to be excluded on the basis of these criteria. All procedures were approved by the Institutional Review Board (IRB) of the University of Connecticut and participants were compensated with a gift card at the conclusion of the study.

[FIGURE 1]

Experimental Design

The study was conducted over three different test sessions completed for all participants over the span of three weeks in the following order: (1) research lab setting, (2) a home environment, (3) research lab setting once more (Figure 2). All testing was completed by one of two testers. The same single-family house was used for all home testing, all on the same date, such that all 12 participants were tested on the same day (back-to-back, between the hours of 10:00am and 4:30pm). This allowed us to minimize travel and set-up time, streamline the data collection timeline, and restrict the number of variables being compared (e.g., two locations).

[FIGURE 2]

Auditory Brainstem Response (ABR) Protocol

The Bio-logic Navigator Pro AEP system (Natus Medical, Inc.) ran both stimulus delivery and averaging. The same laptop computer, and all associated system hardware, were used in both locations. The system is marketed as being portable and lightweight. For the two research lab test sessions, ABRs were recorded in an electromagnetic-shielded double walled sound booth (IAC Acoustics, Winchester, United Kingdom) following standard practices in the lab, with the experimenter and all hardware located outside the booth, except the transducers and patient cable (interface between the electrodes and amplifier). To minimize muscle movement during the recordings, participants sat comfortably in a reclined chair while watching a self-selected, muted movie with English captions. The movie was projected onto the wall of the booth, about five feet from the participant's head, using a ceiling mounted LCD projector located outside the booth window. All lights were shut off, and power was cut in the booth, as is standard practice for recording ABRs in the lab. Participants were asked to leave phones and other personal devices (e.g. smart watches) outside of the booth.

For the home test session, ABRs were recorded in the living room of a single-family home. The house has an open floor plan design such that the living room is continuous with the kitchen and main entrance. All lights were turned off to minimize electrical noise, though appliances in the home, many within the close vicinity of the testing set-up, remained plugged in (e.g. a television) and/or running (e.g. the refrigerator). Muted movies were played through a laptop placed on a table about three feet in front of the participant. The laptop was not plugged in and ran from battery; however, Wi-Fi was enabled on the computer. Participants were asked to leave all personal electronics on the kitchen table, away from the immediate recording site. Testing equipment was situated about three feet behind the participant and the participant was angled so as to not see the recording computer behind them. For the duration of the home session, participants sat comfortably and relaxed on an upholstered stuffed armchair in the living room. The testers remained quiet and as motionless as possible, sitting behind the participants with the recording equipment during the testing to avoid distracting the participant. The configuration of the simulated home testing environment replicated the lab setting as closely as possible. Outside of moving the small table into place to hold the laptop used to present movies, the home environment was not altered from its typical configuration to maintain the realism of the setting. Because participants were tested back-to-back, the kitchen area served as a waiting room. For all three test sessions, the ABR collection, including electrode application set-up, lasted approximately 20 minutes.

To measure the background sound levels of both environments, the National Institute for Occupational Safety and Health (NIOSH) sound level meter smart phone application utilizing the internal microphone on an iPhone 11 (80 dB threshold level with a 3 dB exchange rate) was used. This phone application, rather than a professional-grade sound level meter, was chosen for

it's easy portability and accessibility for future home-based applications. Measurements were taken post hoc (after the full dataset was collected) and the NIOSH application was calibrated prior to use. The average LAeq over a 30 second period for the laboratory sound booth was 29.9 dBA (maximum level = 41.4 dBA) and 28.8 dBA (maximum level = 31.4 dBA) for the home. The home measurement was taken at a particularly quiet time. The phone was placed by the chair, near where the participant's right ear would be, in both locations, to measure sound levels.

Recording Parameters

At each test session, ABRs were recorded to two different stimuli, a click and a /da/ speech stimulus, each repeated isochronously to the right ear in its own testing block (click: 31.3 Hz, 70 dB nHL, 100 μs, rarefaction polarity; /da/: 10.9 Hz, 80 dB SPL, 40 ms, alternating polarity). Recording parameters matched published guidelines. (For a description of the acoustic parameters of the 40 ms /da/ stimulus, refer to Banai et al., 2009.) A three-electrode, conventional, ipsilateral electrode montage (Cz = non-inverting electrode, A2 right earlobe = inverting electrode, forehead = ground electrode) was used for all recordings using gold cup electrodes. The same set of electrodes was used at each session to limit variability between and across participants. Each electrode site was prepped by gentle exfoliation with a cotton swab coated with skin prep gel, followed by the application of electrodes using conductive paste. Impedance for each electrode for all participants was rigorously maintained at 1 kΩ to control for impedance between and across sessions, with frequent re-checks of impedance during each recording session. For the click stimulus, a 100-1500 Hz filter was used with an averaging window extending from 0-10 ms. A 100-2000 Hz filter was employed for recordings to the /da/ stimulus, with a recording window that began -16.2 ms prior to the stimulus onset and 29.13 ms post stimulus offset for a total recording window of 84.56 ms. The click stimulus included 256

sample points and the /da/ stimulus 1024 sample points. The amplifier gain was set to 100,000 for both stimuli. During each session, two blocks were run for each stimulus. Once 2000 artifact-free trials were reached for each block of the click stimulus, or 3000 artifact-free trials were reached for /da/, the recording automatically ended. Bio-logic AEP's default artifact rejection criterion for ABRs (±23.8 μV) was selected, resulting in any trial exceeding that range being rejected from the average. The total number of rejected trials for each run was also logged for analysis. While the convention for our lab is to re-run any block with greater than a 10% rejection rate, this did not need to occur for any participants at any session. The average waveform from the two blocks was manually created and saved offline for each stimulus.

Statistical Analyses

We focus our analyses on three response indices: (1) absolute latencies (reported in ms) from recordings of the click (waves I, III, and V) and /da/ (waves V, A, D, E, F, and O) stimuli, (2) RC to the /da/ stimulus (over the 19.5-42.2 ms time window, encompassing waves D, E, and F) reported as correlation coefficient (r-value); and (3) SNR reported as a ratio. The SNR is the quotient of the quadratic mean of the post-stimulus period (19.5-44.2 ms) divided by the quadratic mean of the pre-stimulus period (-16.2-0 ms). An SNR of < 1.5 is considered unfavorable, and the lab convention is to exclude from analyses, though this did not apply to any of the recordings in the current study.

Wave peak picking was manually completed by a trained member of our research team, and reviewed by a second member, and then confirmed by a third expert rater. RC to the speech-ABR, which has gone by the name of "response stability" or "within-session repeatability" in some publications, was computed in line with standard practice for this stimulus and recording system (Hornickel & Kraus, 2013; Skoe et al., 2015; Tecoulesco et al., 2020). (Specifically, to

calculate the RC, the average waveform of each block of the /da/ stimulus was treated as a time-series and used to find the linear correlation between blocks, using the Pearson's product moment correlation function.) R-values closer to 1 indicate greater morphological consistency between the two blocks. Due to the non-normal distribution of the sampling distribution of r-values, a Fisher z-transformation (z' = 0.5 *ln (1+r/1-r) was applied to the RC correlation coefficients to normalize the distribution. The Fisher z-transformation scores (z') are used in the analyses and in scatter plots. All ABR data analyses, including statistical analyses, were completed using custom routines implemented in MATLAB (The MathWorks, Inc., Natick, MA).

Descriptive statistics (e.g. mean, range) are reported for all three components at each session. Furthermore, comparisons across all three sessions, and between locations (Session 1 vs. Session 2 and Sessions 2 vs. Session 3) and within location (Session 1 vs. Session 3), were quantified for each dependent variable (latency, RC, SNR) by using a set of statistical methods:

1. Pearson's correlations were conducted to examine the relation between all combinations of sessions. Correlations allowed us to examine the within and between location relations, in order to help better understand the effect of location, specifically. Pearson's correlations have been used to examine test-retest reliability in the field of hearing sciences (e.g. Fournier & Héber, 2013; Ku et al., 2015). It is a common stance that test–retest correlations above r = 0.80 are considered good, while those below r = 0.70 are unacceptable (Cicchetti, 1994).
2. Intraclass correlations (ICCs) using a two-way mixed model evaluating the absolute agreement were calculated to compare the response indices in terms of their repeatability between locations (lab, home). Similar to the Pearson's correlation, the

ICC estimates the magnitude of a relation between variables, however, it can also account for differences in the means and be utilized in cases, such as this study, where there more than two timepoints of measurement (Liu et al., 2016). Strong ICCs, between 0.75 and 0.9, and those greater than 0.90, suggest "good" or "excellent" reliability, respectively (Koo & Li, 2016). Here, for each wave, we calculated the magnitude of the relation overall across all three test sessions, as well as pair-wise comparisons between and within test locations. ICC has been used in the AEP literature in examining test-retest reliability of both cortical and subcortical potentials (e.g., Tremblay et al., 2003; Rentzsch et al., 2008; Bidelman et al., 2018; Prendergast et al., 2018).

3. Linear mixed effects modelling was used to assess differences between locations on the ABR with location as a fixed factor and participant ID as a random factor to control for subject. Each dependent variable was treated in a separate analysis. Restricted maximum likelihood estimations were conducted, and the mixed model included random intercepts to take the inter-subject variability. This proved to be the best fit model using the "smaller is better" criteria based on the Akaike's Information Criterion (AIC) (Gurka, 2006). Similar mixed models have been used in existing AEP test-retest literature (e.g., Bidelman et al., 2018).

RESULTS

ABR Wave Latencies

We measured suprathreshold click-evoked and speech-ABRs across three test sessions in two different recording locations (in a research lab at Session 1 and Session 3, and within a home for Session 2). Our first set of analyses focused on ABR wave absolute latencies—the time from the stimulus onset to the peak of the wave. Descriptive statistics are provided for both click

(waves I, III, and V) and /da/ (waves V, A, D, E, F, and O) recordings for all three sessions (Table 1).

[TABLE 1]

Group average waveforms are shown for the click and speech stimulus in Figure 3, along with the waveforms from a representative participant. Overall, latencies show high repeatability between locations (research lab and home). For each wave, we first conducted Pearson's correlations to examine the relations between sessions (Table 2). For the click stimulus, Pearson's correlations showed strong positive relations ($r > 0.7$, $p < 0.007$ for wave I; $r > 0.9$, $p < 0.000$ for wave III; $r > 0.8$, $p < 0.000$ for wave V) across all pair-wise comparisons for the latencies. Relations for all waves from the /da/ stimulus pattern similarly, with strong positive correlations ($r > 0.8$, $p \leq 0.001$) for waves V, A, D, E, and F. Wave O also shows a significant correlation between Session 1 and 2 ($r = 0.832$, $p < 0.001$); the other comparisons for wave O are still significant but have lower r-values (Sessions 1 and 3: $r = 0.639$, $p = 0.025$, Sessions 2 and 3 ($r = 0.586$, $p = 0.045$)). Figure 4 plots the relations between sessions for wave V, both for the click and /da/ stimuli. Wave V is highlighted in plots as it is the most robust component found in a human ABR (Hall, 2007) and it is common across both stimuli. Likewise, Figure 5 uses line plots to illustrate the latency for wave V across sessions, both as a grand average and by participant.

[FIGURE 3]

[TABLE 2]

[FIGURE 4]

[FIGURE 5]

For the ICC, we measured the reliability across all three sessions (Session 1, Session 2, and Session 3), between locations (Session 1 vs. Session 2, Session 2 vs. Session 3), and within location (Session 1 vs. Session 3). ICCs were significant for all waves, with ICCs > 0.7 (Table 3). Figure 6 highlights ICCs for click and speech-ABR wave V latencies.

[TABLE 3]

[FIGURE 6]

Finally, we evaluated whether for each wave the latency differed between test locations via a mixed model. Latency was the dependent variable, with location included in the model as a fixed factor, and participant ID as a random factor. We found no significant differences between locations for any of the nine waves (Table 4).

[TABLE 4]

Response Consistency (RC) and Signal-to-Noise Ratio (SNR)

Descriptive statistics for RC and SNR can be found in Table 5. The mean RCs at each location are in the range of r = 0.8 (see Skoe et al., 2015 for normative data on RC and latencies), and SNRs approximately 4, on average. Collectively, this indicates that the group of recordings generally has a robust SNR and that blocks are strong matches to each other.

[TABLE 5]

Using Pearson's correlations (Figure 7), relations are strong and significant ($r > 0.7$, $p < 0.005$) when calculated between both lab sessions (Session 1 vs. Session 3) for both measures. Weak, non-significant correlations ($r < 0.3$, $p > 0.3$) were found between the lab vs. home settings (Session 1 vs. Session 2, and Session 2 vs. Session 3) for both RC and SNR. (All correlations are positive.) Further, when computing ICCs, the reliability of the RC and SNR are "good" within the lab location but are considered "poor" to "moderate" when calculated between

the home and lab locations, and “moderate” when computing the overall ICC across all three sessions (Table 6; Figure 8).

These reliability estimates are influenced by two participants who show discrepant RC results from session to session. In one participant, RC values, from Session 1 through Session 3, varied from r = 0.79, to 0.50, to 0.92. For the other participant, RC values went from r = 0.44, to 0.93, to 0.71. Other participants were stable between r = 0.8 and 0.9. When removing these two participants from the analysis and repeating ICC calculations, ICCs improve for all between-location calculations (see Table 6). Importantly, for both participants, the home environment did not produce the lowest (i.e., worst) values of their three sessions.

[FIGURE 7]
[FIGURE 8]
[TABLE 6]

Next, we conducted separate mixed models to examine any consistent differences between locations with RC and SNR as dependent variables. No appreciable differences were found between locations for either RC ($F(1,34) = 0.001$, $p = 0.979$) or SNR ($F(1,34) = 0.0162$, $p = 0.900$).

These results motivated a series of follow-up analyses to examine what might contribute to the home and lab recordings not being statistically different between locations but at the same time more variable between location compared to latencies. Differences in artifact level were considered first, as higher levels of artifact can impact wave morphology and decrease SNR, even when the artifact count falls within acceptable limits (Maruthy et al., 2015). Raw number of rejected artifacts was logged for both the click and /da/ recordings, with the number of rejects summed from both blocks of each stimulus type separately. Initial analyses involved paired samples t-tests between all session combinations for both the click and speech-ABR artifacts,

with no significant differences between means found ($p > 0.160$), and so the variable was dropped from subsequent analyses. Although artifact numbers did not differ on average between locations or sessions, we note that one of the two participants mentioned above as being an outlier influencing the RC correlation had the highest number of artifacts compared to any other participant, and that this was the case for all three sessions.

Finally, similar patterning between RC and SNR led us to suspect a relation between the two response indices, that is that variation in SNR across sessions could be driving the RC findings. To examine this, we first conducted Pearson's correlations to describe the relation between RC and SNR. Upon detecting a significant relationship between the two when pooling data of all three sessions ($r = 0.720$, $p < 0.000$), we also evaluated the relationship between RC and the two subcomponents of the SNR calculation (the quadratic mean of the post-stimulus period and the quadratic mean of the pre-stimulus period) to determine whether one or both components were influencing the relation between RC and SNR (i.e., Were higher RCs associated with lower pre-stimulus activity, larger FFRs, or both?). Figure 9 shows a significant positive correlation between RC and SNR, and between RC and post-stimulus RMS ($r = 0.428$, $p = 0.009$), such that higher RC maps on to higher SNRs and larger FFR amplitudes. RMS amplitude of the pre-stimulus activity shows a negative correlation with RC ($r = -0.517$, $p = 0.001$). That is, lower nonstimulus activity maps on to higher RC. However, as can be seen in Figure 9, there are no clear patterns between location and any of the SNR variables.

[FIGURE 9]

DISCUSSION

In the present study, we compared click and speech-ABRs collected in a home setting to ABRs collected in a research laboratory. Our findings suggest that ABR latencies have strong

repeatability between environments, and although RC and SNR showed comparatively lower repeatability between the lab and the home, this is not a consequence of the recordings being of poorer quality in the home. This small study adds to the growing movement for both portable EEG testing and in-home healthcare, and improved accessibility for individuals of all ages and abilities. We discuss our findings within the context of previous work, address limitations, and make recommendations to guide future efforts to record ABRs in non-traditional environments.

ABR latencies are of interest for audiologists due to their high reliability in audiologically-normal ears, and their sensitivity to pathologies of the peripheral and/or central auditory system via delays, aberrations, or a complete absence of the wave (Hall, 2007). Therefore, clinically, latency is a useful and important dependent variable to be considered here. All absolute latencies for all waves, both click and speech-ABRs, at all session-to-session comparisons, showed strong Pearson's correlations and ICCs, and none of the waves were significantly different between locations (lab vs. home) in our mixed models. This finding is in line with other reports of high reliability for ABR latency when variables such as tester (e.g. Cobb & Stuart, 2014) or test operator location (e.g. Towers et al., 2005) were manipulated.

Home environments are less acoustically and electrically controlled than sound booths found in a research lab or a clinic with more visual distractions. Therefore, if differences were to have emerged between home and lab settings, we would have expected to see higher levels of motion artifact from distraction, increased difficulty in identifying waves due to contamination from electrical and myogenic artifacts, lower SNRs, and/or delayed ABRs due to higher levels of background noise (Burkard & Sims, 2002; Parbery-Clark et al., 2009; Anderson et al., 2010). However, no significant differences are found between locations.

Next, we examined RC of the /da/ stimulus—a metric reflecting within-session consistency. Lower RC has been noted in children with autism spectrum disorders (Otto-Meyer et al., 2018) and developmental reading difficulties (Hornickel & Kraus), with our group's recent work showing an association with phonological ability in school-age children (Tecoulesco et al., 2020). By contrast, higher response consistency has been noted in musicians (Skoe & Kraus, 2013) and bilinguals (Krizman et al., 2015a). Moreover, RC has been studied in individuals throughout the lifespan, where RC has been shown to increase throughout early childhood (Callaway & Halliday, 1973) and decrease during adolescence (Krizman et al., 2015b).

Investigations into the test-retest reliability of RC are an important step if clinical translation of the measure is to take place. To date, only one previous study has investigated this. Hornickel et al., 2012a found significant reliability of RC between sessions in typically developing children with two measurements, one year apart. Our findings in a smaller dataset show high RC overall, though RC was more variable between locations than latencies were. However, our RC values were consistent with those found in Hornickel et al., 2012a. Pearson's correlations showed stronger relations between the two lab sessions than they did between home and lab sessions. Additionally, ICC was considered "strong" only between the two lab sessions. However, although more variable from location to location, the home environment did not yield lower RCs on average compared to the lab. In six out of 12 participants (50%), their RC was higher in the home session than either lab session, and our mixed model did not show consistent differences in RC between locations. This suggests that in-home recording environments themselves do not necessarily drive poorer within-session consistency in responses, and that the effect of location does not produce a common outcome on everyone's recordings. Our findings in this small dataset encourage follow-up studies into the reliability of RC in different

environments that also factor in other relevant biological and behavioral measures previously shown to correlate with RC, such as phonological discrimination and awareness (e.g. Hornickel et al., 2012b; Tecoulesco et al., 2020).

We also investigated SNR, a measure of response magnitude that takes into account non-stimulus evoked neural activity during the pre-stimulus time window. We found that SNR patterned similarly to RC, also showing some variability between the home and lab sessions—more variability than between the two lab sessions, as assessed by Pearson's correlations and ICCs, but also comparable in its outcome from a mixed model, where no appreciable, consistent differences between locations are found. This motivated us to examine the individual components from which the SNR is derived in order to get insight into whether differences between the home and lab environment were driven by increased noise in the home environment from potentially a variety of sources, including increased electrical noise, increased myogenic, or background neurologic activity, and/or increased acoustic noise during testing. We anticipated that the pre-stimulus period might be correlated with RC given previous work showing age-related changes to RC that inversely parallel changes to pre-stimulus activity (Skoe et al., 2015), such that when the pre-stimulus RMS amplitude was low, the RC was high, and vice versa. While for the current study, the sample is limited to one age group, we still found these two ABR measures coupling together. Our results showed that decreased pre-stimulus activity and increased stimulus-related activity were both linked to higher RC. Decreased pre-stimulus activity may be indicative of a state of lower neural noise, which could foster higher RC by promoting more consistent responses across trials and translate to increased stimulus-related amplitudes. Indeed, work on animal models show that low neural stability reflects increased single-trial variation in intracellular brainstem activity (White-Schwoch, Nicol, Warrier, Abrams,

& Kraus., 2017). Decreased pre-stimulus activity could also be indicative of lower electrical noise or lower myogenic noise, which would enable the response to be recorded at a higher SNR and allow for the response to appear more consistent.

Similar to RC, SNR was not lower (i.e. poorer) in the home, on average, than in the lab. (See Table 5 for means.) (Of importance—artifact rejection rates did not differ between lab and home sessions.) Combining across all subjects and sessions, there was a mean SNR of 4.181 (SD = 1.567). Because only a SNR of less than 1.5 is considered "unfavorable" (Skoe & Kraus, 2010), we did not have to recollect any data or exclude any subjects either in the home or in the lab.

These outcomes of our analysis of the RC suggest that, while the home testing environment itself does not drive changes in RC, the SNR of the recording, which can differ from session to session regardless of location, does influence RC, and that a lower SNR can worsen the overall morphological consistency between two blocks of speech-ABR recordings within a session. Considerations for maximizing SNR should be made when possible; measures to keep the patient relaxed and comfortable (e.g. providing a reclining chair, instructions to try to relax and not to tense up) should be implemented in any test location to keep muscle activity to a minimum. From experience, we know that muscle tension can influence the quality of the ABR even when the participant appears to be sitting still with no movements. Because we did not stringently control for acoustic noise from outside sources that could have varied throughout the home session (e.g. unplugging all electrical devices except the test equipment, turning off Wi-Fi, removing all other people and pets from the home), and sound level measurements were not made throughout the test session, the variability between locations could reflect fluctuations in the home environment that do not occur in the lab (e.g., cycling on and off major electrical

appliances, intermittent noise from other people or pets in the room). Future work could investigate this possibility by making serial recordings in the same individual in both locations.

In-home ABR testing, and portable EEG more broadly, is not without its limitations, nor is the current study. Unlike research labs or clinics, homes and remote locations are largely uncontrolled and can be unpredictable in many ways. In the current study, we used a simulated home environment, where all participants were tested in one common home. As shown with the sound level measurements, our home environment had an acoustic environment on-par with that of the sound booth within the lab setting, though the measurement was taken at home at a particularly quiet moment. Still, we collected our ABR data at a time when there was limited outside traffic or other activities in the home. This will not always in the case in all test homes. Configurations of homes differ, with some potentially less suitable for in-home electrobiological equipment and measurements than others. Also, though we did not find electrical noise (e.g. 60 Hz line noise) to be a problem our test home, it may prove to be more problematic in others. Additionally, as noted, noise created by the surrounding environment (e.g. other family members, neighbors, pets, busy streets) is a concern that exists in a home, but not in a sound booth within a lab or clinic. However, again, while we did record all in-home ABRs in one single-family home, we purposely left the home and surroundings operating as naturalistically as possible to be representational of any given home.

Additionally, purchasing ABR equipment that is specifically created and widely marketed to be portable (e.g. Wiegers et al., 2015) can be expensive, especially for labs or clinics that already own and are using a non-portable system. The current study demonstrates that some existing clinical ABR systems (e.g. the Bio-logic Navigator Pro AEP) can easily be easily transportable with a laptop computer and brought into the home.

Regarding the current study, our sample size is limited to 12 participants, all young adults with normal hearing. This small size may have contributed to the findings for RC and SNR given that two participants appeared to influence the ICC and correlation analysis. Moreover, the majority of our participant group (but not all) was already familiar with ABRs or had sat for one previously. A larger, more heterogeneous participant group, in terms of age, hearing status, or experience with ABRs, could help further examine the utility of in-home recordings in clinical practice. However, participant selection was not entirely unintentional, as we wanted to minimize additional confounding variables and maximize the advantage of having participants with experience sitting for this type of investigation, where having relaxed participants mitigated any effects due to test familiarity. Investigations into clinical reliability for in-home ABRs should include a larger sample size.

A study design limitation is that we did not include a second session in the home to examine the repeatability within two home sessions. Additionally, while we limited our in-home testing to one single house, testing in a wider variety of homes will allow for a broader interpretation of how different locations might affect ABR recordings.

CONCLUSIONS

Our results show that ABR latencies, both for click-evoked and speech-ABRs, are highly comparable between common lab and home settings. RC and SNR, though more variable between locations, were, on-average, as consistent and clean (i.e. high) in the home as they were when recorded in the lab, and the two measures couple together, where higher SNR can drive higher RC. Reliable portable ABR testing can be valuable for providing home healthcare services, potentially making audiological services more available to a variety of populations, including those who are lacking transportation options, are difficult to test, have disabilities, or

live in remote areas. Additionally, such portable equipment opens doors for bringing EEG research into the field and expanding participant samples both in size and diversity. The current study adds to the emergent movement both for home healthcare in audiology and portable EEG testing across research fields and shows that high-quality, reliable recordings can be obtained outside of a sound booth using standard equipment.

Acknowledgements

We wish to thank Letitia Naigles and Vivi Tecoulesco for their helpful discussions relating to study design.

Statement of author contributions

A.P. and C.S. performed the experiments. A.P. analyzed the data and wrote the paper. C.S. contributed to the introduction, methods, and discussion sections of the paper. E.S. designed the experiments and provided critical feedback and editing on the manuscript at all stages.

References

Achor, L. J., & Starr, A. (1980). Auditory brain stem responses in the cat. II. Effects of lesions. *Electroencephalography and Clinical Neurophysiology*, *48*(2), 174-190.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *Journal of Neuroscience*, *30*(14), 4922-4926.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience*, *32*(41), 14156-14164.

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and subcortical auditory function. *Cerebral Cortex*, *19*(11), 2699-2707.

Bennett, S., Laver, K., Voigt-Radloff, S., Letts, L., Clemson, L., Graff, M., ... & Gitlin, L. (2019). Occupational therapy for people with dementia and their family carers provided at home: a systematic review and meta-analysis. *BMJ Open*, *9*(11).

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage*, *175*, 56-69.

Bidelman, G. M., Pousson, M., Dugas, C., & Fehrenbach, A. (2018). Test-Retest Reliability of Dual-Recorded Brainstem versus Cortical Auditory-Evoked Potentials to Speech. *Journal of the American Academy of Audiology*, *29*(2), 164-174.

Bitterman, N. (2011). Design of medical devices—A home perspective. *European Journal of Internal Medicine*, *22*(1), 39-42.

Burkard, R. F., & Sims, D. (2002). A comparison of the effects of broadband masking noise on the auditory brainstem response in young and older adults. *American Journal of Audiology*.

Callaway, E., & Halliday, R. A. (1973). Evoked potential variability: Effects of age, amplitude and methods of measurement. *Electroencephalography and clinical neurophysiology*, *34*(2), 125-133.

Calyam, P., Mishra, A., Antequera, R. B., Chemodanov, D., Berryman, A., Zhu, K., ... & Skubic, M. (2016). Synchronous big data analytics for personalized and remote physical therapy. *Pervasive and Mobile Computing*, *28*, 3-20.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284.

Cobb, K. M., & Stuart, A. (2014). Test-retest reliability of auditory brainstem responses to chirp stimuli in newborns. *International Journal of Audiology*, *53*(11), 829-835.

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., ... & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 1-10.

Daspit, C. P., Raudzens, P. A., & Shetter, A. G. (1982). Monitoring of intraoperative auditory brain stem responses. *Otolaryngology--Head and Neck Surgery*, *90*(1), 108-116.

Debener, S., Emkes, R., De Vos, M., & Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific Reports*, *5*, 16743.

Dzulkarnain, A. A. A., Buyong, A. S., & Sulaiman, N. H. (2014). Intra-subject variability in the auditory brainstem response using a vertical montage recording. *Speech, Language and Hearing*, *17*(3), 160-167.

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. *Electroencephalography and Clinical Neurophysiology*, *53*(2), 125-132.

Emblin, K., Nash, I., & Jefferies, S. (2016). Pharmacy at home: service for frail older patients demonstrates medicines risk reduction and admission avoidance. *Medicines Management*, *8*(7), 217-222.

Fournier, P., & Hébert, S. (2013). Gap detection deficits in humans with tinnitus as assessed with the acoustic startle paradigm: does tinnitus fill in the gap?. *Hearing Research*, *295*, 16-23.

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*, *6*(17), 2363-2367.

Galbraith, G. C., Amaya, E. M., de Rivera, J. M. D., Donan, N. M., Duong, M. T., Hsu, J. N., ... & Tsang, L. P. (2004). Brain stem evoked response to forward and reversed speech in humans. *Neuroreport*, *15*(13), 2057-2060.

Giménez-Díez, D., Maldonado Alía, R., Rodríguez Jiménez, S., Granel, N., Torrent Solà, L., & Bernabeu-Tamayo, M. D. (2020). Treating mental health crises at home: Patient satisfaction with home nursing care. *Journal of Psychiatric and Mental Health Nursing*, *27*(3), 246-257.

Godlove, J., Anantha, V., Advani, M., Des Roches, C. A., & Kiran, S. (2019). Comparison of therapy practice at home and in the clinic: A retrospective analysis of the Constant Therapy platform data set. *Frontiers in Neurology*, *10*, 140.

Gorga, M. P., Worthington, D. W., Reiland, J. K., Beauchaine, K. A., & Goldgar, D. E. (1985).

Some comparisons between auditory brain stem response thresholds, latencies, and the pure-tone audiogram. *Ear and Hearing*, *6*(2), 105-112.

Greenberg, S. (1980). *Temporal neural coding of pitch and vowel quality*. University of California.

Grose, J. H., Buss, E., & Hall III, J. W. (2017). Loud music exposure and cochlear synaptopathy in young adults: Isolated auditory brainstem response effects but no perceptual consequences. *Trends in Hearing*, 21, 1-18.

Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician*, 60(1), 19-26.

Hall, J. W. (2007). *New handbook of auditory evoked responses*. Pearson.

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

Hornickel, J., Knowles, E., & Kraus, N. (2012a). Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. *Hearing Research*, *284*(1-2), 52-58.

Hornickel, J., Zecker, S. G., Bradlow, A. R., & Kraus, N. (2012b). Assistive listening devices drive neuroplasticity in children with dyslexia. *Proceedings of the National Academy of Sciences*, *109*(41), 16731-16736.

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: a biological marker of dyslexia. *Journal of Neuroscience*, *33*(8), 3500-3504.

Jafari, Z., & Malayeri, S. (2014). Effects of congenital blindness on the subcortical representation of speech cues. *Neuroscience*, *258*, 401-409.

Johnson, J. L., White, K. R., Widen, J. E., Gravel, J. S., James, M., Kennalley, T., ... & Weirather, Y. (2005). A multicenter evaluation of how many infants with permanent hearing loss pass a two-stage otoacoustic emissions/automated auditory brainstem response newborn hearing screening protocol. *Pediatrics*, *116*(3), 663-672.

Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *Journal of Neuroscience*, *28*(15), 4000-4007.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., & White-Schwoch, T. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *Journal of Neuroscience*, *34*(36), 11913-11918.

Kraus, N., Thompson, E. C., Krizman, J., Cook, K., White-Schwoch, T., & LaBella, C. R. (2016). Auditory biological marker of concussion in children. *Scientific Reports*, *6*, 39009.

Kraus, N., Anderson, S., & White-Schwoch, T. (2017). The frequency-following response: A window into human communication. In *The frequency-following response* (pp. 1-15). Springer, Cham.

Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in Neuroscience*, *11*, 109.

Krizman, J., Slater, J., Skoe, E., Marian, V., & Kraus, N. (2015a). Neural processing of speech in children is influenced by extent of bilingual experience. *Neuroscience Letters*, *585*, 48-53.

Krizman, J., Tierney, A., Fitzroy, A. B., Skoe, E., Amar, J., & Kraus, N. (2015b). Continued maturation of auditory brainstem function during adolescence: A longitudinal approach. *Clinical Neurophysiology*, *126*(12), 2348-2355.

Ku, Y., Ahn, J. W., Kwon, C., Suh, M. W., Lee, J. H., Oh, S. H., & Kim, H. C. (2015). Gap prepulse inhibition of the auditory late response in healthy subjects. *Psychophysiology*, *52*(11), 1511-1519.

Kuziek, J. W., Shienh, A., & Mathewson, K. E. (2017). Transitioning EEG experiments away from the laboratory using a Raspberry Pi 2. *Journal of neuroscience methods*, *277*, 75-82.

Kuziek, J. W., Redman, E. X., Splinter, G. D., & Mathewson, K. E. (2018). Increasing the mobility of EEG data collection using a Latte Panda computer. *Journal of Neuroscience Methods*, *308*, 34-47.

Lau-Zhu, A., Lau, M. P., & McLoughlin, G. (2019). Mobile EEG in research on neurodevelopmental disorders: Opportunities and challenges. *Developmental Cognitive Neuroscience*, 100635.

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., & Tu, X. M. (2016). Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Archives of Psychiatry*, *28*(2), 115–120.

Malayeri, S., Lotfi, Y., Moossavi, S. A., Rostami, R., & Faghihzadeh, S. (2014). Brainstem response to speech and non-speech stimuli in children with learning problems. *Hearing Research*, *313*, 75-82.

Maruthy, S., Gnanateja, G., Ramachandran, R., & Thuvassery, P. (2015). Characterizing Muscle Artifact Interference in AEP Recording. *Journal of Hearing Science*, *5*(3).

McFarland, D. J., & Cacace, A. T. (2011). Covariance is the proper measure of test–retest reliability. *Clinical Neurophysiology*, *9*(122), 1893.

McFarland, D. J., & Cacace, A. T. (2012). Questionable reliability of the speech-evoked auditory brainstem response (sABR) in typically-developing children. *Hearing Research*, *287*(1-2), 1-2.

Mehraei, G., Hickox, A. E., Bharadwaj, H. M., Goldberg, H., Verhulst, S., Liberman, M. C., & Shinn-Cunningham, B. G. (2016). Auditory brainstem response latency in noise as a marker of cochlear synaptopathy. *Journal of Neuroscience*, *36*(13), 3755-3764.

Neef, N. E., Müller, B., Liebig, J., Schaadt, G., Grigutsch, M., Gunter, T. C., ... & Kraus, N. (2017). Dyslexia risk gene relates to representation of sound in the auditory brainstem. *Developmental Cognitive Neuroscience*, *24*, 63-71.

Norton, S. J., Gorga, M. P., Widen, J. E., Folsom, R. C., Sininger, Y., Cone-Wesson, B., ... & Fletcher, K. (2000). Identification of neonatal hearing impairment: evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance. *Ear and Hearing*, *21*(5), 508-528.

Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research*, *236*(3), 733-743.

Oyler, R. F., Lauter, J. L., & Matkin, N. D. (1991). Intrasubject variability in the absolute latency of the auditory brainstem response. *Journal of the American Academy of Audiology*, *2*(4), 206-213.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience*, *29*(45), 14100-14107.

Prendergast, G., Tu, W., Guest, H., Millman, R. E., Kluk, K., Couth, S., ... & Plack, C. J. (2018). Supra-threshold auditory brainstem response amplitudes in humans: Test-retest reliability, electrode montage and noise exposure. *Hearing Research*, *364*, 38-47.

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: a normative dataset in healthy newborns. *Hearing Research*, *371*, 28-39.

Rentzsch, J., Jockers-Scherübl, M. C., Boutros, N. N., & Gallinat, J. (2008). Test–retest reliability of P50, N100 and P200 auditory sensory gating in healthy subjects. *International Journal of Psychophysiology*, *67*(2), 81-90.

Scanlon, J. E., Townsend, K. A., Cormier, D. L., Kuziek, J. W., & Mathewson, K. E. (2019).

Taking off the training wheels: Measuring auditory P3 during outdoor cycling using an active wet EEG system. *Brain Research*, *1716*, 50-61.

Schwartz, D. M., Morris, M. D., & Jacobson, J. T. (1994). The normal auditory brainstem response and its variants. *Principles and Applications in Auditory Evoked Potentials*, 123-153.

Sininger, Y. S., Abdala, C., & Cone-Wesson, B. (1997). Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. *Hearing Research*, *104*(1-2), 27-38.

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing*, *31*(3), 302.

Skoe, E., & Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Frontiers in Psychology*, *4*, 622.

Skoe, E., Krizman, J., & Kraus, N. (2013). The impoverished brain: disparities in maternal education affect the neural response to sound. *Journal of Neuroscience*, *33*(44), 17221-17231.

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*(6), 1415-1426.

Skoe, E., Brody, L., & Theodore, R. M. (2017). Reading ability reflects individual differences in auditory brainstem function, even into adulthood. *Brain and Language*, *164*, 25-31.

Song, J. H., Nicol, T., & Kraus, N. (2011). Test–retest reliability of the speech-evoked auditory brainstem response. *Clinical Neurophysiology*, *122*(2), 346-355.

Tecoulesco, L., Skoe, E., & Naigles, L. R. (2020). Phonetic discrimination mediates the relationship between auditory brainstem response stability and syntactic performance. *Brain and Language*, *208*, 104810.

Toppenberg, M. D., Christiansen, T. E. M., Rasmussen, F., Nielsen, C. P., & Damsgaard, E. M. (2019). Mobile X-ray outside the Hospital: A review.

Towers, A. D., Pisa, J., Froelich, T. M., & Krumm, M. (2005). The reliability of click-evoked and frequency-specific auditory brainstem response testing using telehealth technology. *Seminars in Hearing*, 26(1), 26-34.

Tremblay, K. L., Friesen, L., Martin, B. A., & Wright, R. (2003). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear and Hearing*, *24*(3), 225-232.

Tusa, R. J., Stewart, W. F., Shechter, A. L., Simon, D., & Liberman, J. N. (1994). Longitudinal

study of brainstem auditory evoked responses in 87 normal human subjects. *Neurology*, *44*(3 Part 1), 528-528.

Van der Drift, J. F. C., Brocaar, M. P., & Van Zanten, G. A. (1987). The Relation between the Pure-Tone Audiogram and the Click Auditory Brainstem Response Threshold in Cochlear Hearing Loss. *Audiology*, *26*(1), 1-10.

Vander Werff, K. R., & Burns, K. S. (2011). Brain stem responses to speech in younger and older adults. *Ear and Hearing*, *32*(2), 168-180.

White-Schwoch, T., Nicol, T., Warrier, C. M., Abrams, D. A., & Kraus, N. (2017). Individual differences in human auditory processing: insights from single-trial auditory midbrain activity in an animal model. *Cerebral Cortex*, *27*(11), 5095-5115.

Wiegers, J. S., Bielefeld, E. C., & Whitelaw, G. M. (2015). Utility of the Vivosonic Integrity™ auditory brainstem response system as a hearing screening device for difficult-to-test children. *International Journal of Audiology*, *54*(4), 282-288.

Zink, R., Hunyadi, B., Van Huffel, S., & De Vos, M. (2016). Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *Journal of Neural Engineering*, *13*(4), 046017.

FIGURE CAPTIONS

Figure 1. Average hearing level thresholds across participants prior to Session 1. All participants were required to have clinically normal hearing at all test frequencies in both ears. Error bars represent the standard error of the mean.

Figure 2. Study timeline. Three test session occurred over the span of three weeks. Session 1 and Session 3 occurred in a research lab setting and Session 2 occurred in a home.

Figure 3. Grand-average (left) ABR waveforms and waveform from one individual representative participant (right) for the click and /da/ stimuli comparing Session 1 (lab, plotted in grey), Session 2 (home, plotted in purple for the click and red for the /da/ stimulus), and Session 3 (lab, plotted in black). For both stimuli, all show high reliability across all test sessions.

Figure 4. Scatter plots of the relation between Wave V latencies for the click (top, purple) and /da/ (bottom, red) stimuli, between Session 1 (lab), 2 (home), and 3 (lab). Pearson's correlations show a strong relation between sessions at all comparisons for wave V latency for both stimuli. For the /da/ stimulus, latencies for three participants measured to be identical (6.55 ms), thus overlap on the plot.

Figure 5. Line plots of Click Wave V latency (top) and /da/ Wave V latency (bottom). In the left column, the line in each plot represents the average across participants, with error bars representing the standard error of the mean. In the right column, each line represents one individual participant.

Figure 6. Intraclass correlations for Click Wave V latency (purple) and /da/ Wave V latency (red). In the first group (left), ICCs are calculated across all three sessions (Session 1—lab; Session 2—home; Session 3—lab). In the second group, ICCs are calculated between lab and home locations (Sessions 1 and 2). In the third group, ICCs are calculated within the lab location (Sessions 1 and 3), and in the fourth group, between locations again (Sessions 2 and 3). Strong ICCs are found for click and /da/ wave V latencies across the board. Descriptive labels (poor, moderate, good, and excellent) are adopted from standard conventions for the ICC (Koo and Li, 2016).
** = $p < 0.000$

Figure 7. Scatter plots of the relation between response consistency (z' scores—top), and signal-to-noise ratio (bottom) between Sessions 1 (lab), 2 (home) and 3 (lab). Pearson's correlations show strong correlations between in-lab sessions, but weak correlations between lab and home for both response indices.

Figure 8. Intraclass correlations for response consistency (RC—blue) and signal-to-noise ratio (SNR—yellow). In the first group (left), ICCs are calculated across all three sessions (Session 1—lab; Session 2—home; Session 3—lab). In the second group, ICCs are calculated between lab and home locations (Sessions 1 and 2). In the third group, ICCs are calculated within the lab location (Sessions 1 and 3), and in the fourth group, between locations (Sessions 2 and 3). In the calculation that includes all three sessions (left), ICCs are found to be moderate. Calculations include all participants and do not exclude outliers.
* = $p < 0.05$
** = $p < 0.001$

Figure 9. Pearson's correlations show the strong relations between response consistency (z') signal-to-noise ratio and its subcomponents. Correlations were conducted with data points pooled from all three sessions.
S = Session
L = Lab
H = Home