

## Test–retest reliability of the speech-evoked auditory brainstem response

Judy H. Song<sup>a,b</sup>, Trent Nicol<sup>a,b</sup>, Nina Kraus<sup>a,b,c,d,\*</sup>

<sup>a</sup> Auditory Neuroscience Laboratory, Northwestern University, Evanston, IL 60208, USA

<sup>b</sup> Department of Communication Studies, Northwestern University, Evanston, IL 60208, USA

<sup>c</sup> Department of Neurobiology and Physiology, Northwestern University, Evanston, IL 60208, USA

<sup>d</sup> Department of Otolaryngology, Northwestern University, Evanston, IL 60208, USA

### ARTICLE INFO

#### Article history:

Accepted 4 July 2010

Available online 16 August 2010

#### Keywords:

ABR test–retest reliability

Speech-evoked ABR

Speech-in-noise

Brainstem encoding

BioMARK

### ABSTRACT

**Objective:** The speech-evoked auditory brainstem response (ABR) provides an objective measure of subcortical encoding of complex acoustic features. However, the intrasubject reliability of this response in both optimal and challenging listening conditions has not yet been systematically documented. This study aimed to evaluate test–retest reliability of the speech-evoked ABR in young adults.

**Methods:** In each of two sessions, ABRs were obtained with: (1) a 170 ms /da/ syllable presented in quiet as well as 2-talker and 6-talker babble background noise conditions and (2) a 40 ms /da/ syllable presented in quiet. Test–retest reliability of the responses was analyzed in the frequency and time domains. **Results:** The speech-evoked ABR does not vary significantly across sessions within individuals on measures of temporal encoding (i.e., peak latencies, stimulus-to-response and response-to-response measures), frequency representation and response magnitude.

**Conclusions:** The subcortical auditory pathway produces a response to a complex sound that is stable and replicable from session to session.

**Significance:** By demonstrating the high degree of replicability in optimal and challenging listening conditions, the applicability of the speech-evoked ABR may be increased to examining a range of auditory processing abilities in clinical and research settings.

© 2010 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

### 1. Introduction

For decades, the highly predictable nature of the click-evoked auditory brainstem response (ABR), demonstrated by its astounding degree of both temporal precision and test–retest reliability, has made it a widely used clinical measure to assess the integrity of the subcortical pathway (Jacobson, 1985; Cone-Wesson et al., 1987, 2000; Hall, 1992; Sininger, 1993; Sininger et al., 1997; Hood, 1998; Norton et al., 2000). It is a far-field response that reflects stimulus-locked, synchronous neural firing from nuclei along the brainstem (Møller and Jannetta, 1985; Chandrasekaran and Kraus, 2010). The intrasubject variability of absolute latencies of the click-evoked ABR from one test to another is small (Edwards et al., 1982; Oyler et al., 1991). Thus, temporal precision of the response is such that even subtle divergences in peak latencies on the order of fractions of milliseconds are clinically significant.

More recently, the ABR has been used to evaluate underlying central processes involved in encoding more complex signals pre-

\* Corresponding author at: Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA. Tel.: +1 847 491 2465; fax: +1 847 491 2523.

E-mail address: [nkraus@northwestern.edu](mailto:nkraus@northwestern.edu) (N. Kraus).

URL: <http://www.brainvolts.northwestern.edu> (N. Kraus).

sented in diverse listening contexts. Among these are syllables that have been temporally and spectrally enhanced (Cunningham et al., 2002), synthesized with Mandarin tones (Krishnan et al., 2004, 2009; Song et al., 2008; Swaminathan et al., 2008) or presented in background noise (Russo et al., 2005; Parbery-Clark et al., 2009; Anderson et al., 2010; Song et al., in press) as well as words (Galbraith et al., 1995, 1997, 2004), musical notes and chords (Musacchia et al., 2007; Lee et al., 2009) and emotionally-valent vocal sounds (Strait et al., 2009). The brainstem response to a consonant–vowel syllable, such as /da/, is composed of two regions: the onset response that signals the beginning of the sound and the frequency following response (FFR) that corresponds to the periodicity of the vowel (Boston and Møller, 1985; Akhoun et al., 2008; Johnson et al., 2008a; Hornickel et al., 2009b). Although the morphology of the onset response of the speech-evoked ABR is similar to that elicited by clicks (Song et al., 2006), the FFR does not appear in the transient click-evoked response. The FFR faithfully reflects the encoding of the fundamental frequency and harmonic structure of the speech stimulus (Moushegian et al., 1973). Thus, an advantage of the speech-evoked ABR is that it reflects both transient and sustained portions of the stimulus which can be objectively assessed at the level of the brainstem.

Investigating the brainstem encoding of more complex sounds is important as responses to these stimuli can uncover auditory processing deficits (Song et al., 2006; Banai et al., 2009) and experience-dependent enhancement (Musacchia et al., 2007; Song et al., 2008; Krishnan et al., 2009; Lee et al., 2009; Parbery-Clark et al., 2009; Strait et al., 2009; Bidelman et al., 2009) of subcortical encoding that clicks alone cannot. Click and speech stimuli entail different subcortical processing demands such that speech stimuli impose larger desynchronizing influences on neural phase-locking. Specifically, unlike the click (a nonperiodic, relatively simple sound that is very brief in duration and contains a broad range of frequencies), consonant–vowel speech syllables, such as the /da/ used in this study, begin with rapid, relatively low amplitude transient onset features that are particularly vulnerable to disruption by background noise (Brandt and Rosen, 1980). The vowel that follows, a sustained periodic acoustic signal that is more intense than the consonant, may mask the brief consonant onset crucial for eliciting the onset portion of the speech-evoked ABR. Susceptibility to neural phase-locking desynchronization due to the acoustics of the signal may in turn impart greater challenges in replicating the response with good reliability from one session to another. This may be especially apparent in suboptimal listening conditions (e.g., background noise) where degradation and variability of the response occur naturally. Establishing the degree of intrasubject response reliability of speech-evoked brainstem responses recorded in quiet and noisy background conditions will enhance clinical and research utility.

While studies that investigated the impact of auditory training programs on subcortical responses have corroborated the reliability of the speech-evoked ABR recorded over multiple sessions, a thorough examination of the stability of the multiple components of these responses was not performed. Russo and colleagues (2005) found training related improvements in the precision of brainstem encoding in noise in children, whereas control subjects showed stable test–retest reliability on most measures of brainstem activity. Also in the context of a training paradigm, Song and colleagues (2008) recorded over multiple sessions brainstem responses elicited by a speech sound that was synthesized with different Mandarin tones in young adults. They examined the degree of accuracy to which the response encoded the fundamental frequency ( $F_0$ ) trajectory of the stimuli. Responses elicited by the control stimuli did not change in  $F_0$  encoding from pre- to post-training test sessions. The study did not include test–retest analysis of harmonic encoding nor measures of timing, and by performing a wider scope of analysis in the temporal and frequency domains of the speech-evoked ABR, the current study aims to provide a broader understanding of the stability and highly predictable nature of the brainstem response in encoding complex sounds.

To determine whether test–retest reliability is a factor when assessing response timing and morphology, we first need to describe the nature and range of the response among normal subjects. This delineation also benefits intrasubject clinical applications of the speech-evoked ABR (e.g., neural index of change following auditory training). The purpose of the present study was to evaluate test–retest reliability of various speech-evoked ABR measures observed in normal hearing young adults. Specifically, we hypothesized that the neural system of a young adult produces a response to a complex sound that is stable and replicable from session to session.

## 2. Methods

### 2.1. Participants

All participants had normal IQ  $\geq 90$  as measured by the Wechsler's Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999) or

Test of Nonverbal Intelligence-3 (TONI-3) (Brown et al., 1997), normal hearing ( $\leq 20$  dB HL pure-tone thresholds from 125 to 8000 Hz), normal click-evoked auditory response wave V latencies to 100  $\mu$ s clicks presented at 31.1 Hz at 80.3 dB SPL, and no history of neurological disease. All participants gave their informed consent in accordance with the Northwestern University Institutional Review Board regulations.

#### 2.1.1. Study 1

Thirty-one young adults (20 females, ages 19–31 years, mean age = 23.5, SD = 3 years) were tested with the 170 ms /da/ in quiet and background noise. The mean time between Tests 1 and 2 was 58 ( $\pm 33$ ) days.

#### 2.1.2. Study 2

Forty-five young adults (29 females, ages 19–36 years, mean age = 24.5, SD = 4 years) were tested with the 40 ms /da/. The mean test time between Tests 1 and 2 was 41 ( $\pm 34$ ) days.

## 2.2. Neurophysiologic stimuli and recording parameters

### 2.2.1. 170 ms /da/, quiet and background noise

Brainstem responses were elicited in response to the syllable /da/ in quiet and two noise conditions in two test sessions separated by a  $\sim 2$  month interval. /da/ is a six-formant syllable synthesized at a 20 kHz sampling rate using a Klatt synthesizer (Klatt, 1980). The duration was 170 ms with voicing (100 Hz fundamental frequency) onset at 10 ms. Formant transition duration was 50 ms and comprised a linearly rising  $F_1$  (400–720 Hz), linearly falling  $F_2$  and  $F_3$  (1700–1240 and 2580–2500 Hz, respectively) and flat  $F_4$  (3300 Hz),  $F_5$  (3750 Hz) and  $F_6$  (4900 Hz). After the transition period, these formant frequencies remained constant at 720, 1240, 2500, 3300, 3750, and 4900 Hz for the remainder of the syllable. The stop burst consisted of 10 ms of initial frication centered at frequencies around  $F_4$  and  $F_5$ . The syllable /da/ was presented in alternating polarities via a magnetically shielded insert earphone placed in the right ear (ER-3, Etymotic Research, Elk Grove Village, IL) at 80.3 dB SPL at a rate of 4.35 Hz.

The noise conditions consisted of multi-speaker babble spoken in English. Two-talker (1 female and 1 male, 20 s track) and 6-talker (3 females and 3 males, 4 s track) babble were selected because they provide different levels of energetic masking. To create the babble, the speakers were instructed to speak in a natural, conversational style. Recordings of nonsense syllables were made in a sound-attenuated booth in the phonetics laboratory of the Department of Linguistics at Northwestern University for unrelated research (Smiljanic and Bradlow, 2005) and were digitized at a sampling rate of 16 kHz with 24 bit accuracy (for further detail, see Smiljanic and Bradlow, 2005; Van Engen and Bradlow, 2007). The tracks were RMS amplitude normalized using Level 16 software (Tice and Carrell, 1998). The 2-talker babble and 6-talker babble tracks were looped for the duration of data collection (approximately 25 min per condition) with no silent intervals. This presentation paradigm allowed the noise to occur at a randomized phase with respect to the target speech sound. Thus, responses that were time-locked to the target sound could be averaged without the confound of phase-coherent responses to the background noise.

In each of the quiet and two noise conditions, 6300 sweeps of /da/ were collected using Scan 4.3 Acquire (Compumedics, Charlotte, NC) in continuous mode at a sampling rate of 20 kHz. The continuous recordings were filtered, artifact rejected ( $\pm 35$   $\mu$ V), and averaged off-line using Scan 4.3. Responses were band-pass filtered from 70 to 1000 Hz, 12 dB/octave. Waveforms were averaged with a time window spanning 40 ms prior to the onset and 16.5 ms after the offset of the stimulus and baseline corrected over the

pre-stimulus interval (–40 to 0 ms, with 0 corresponding to the stimulus onset). Responses of alternating polarity were added to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Gorga et al., 1985). The final average response consisted of the first 6000 artifact free responses.

### 2.2.2. 40 ms /da/

The speech syllable is a 40-ms /da/ syllable synthesized at a sampling rate of 10 kHz using a Klatt synthesizer (Klatt, 1980). It contains a release burst and voiced formant transition with a fundamental frequency ( $F_0$ ) that linearly rises from 103 to 125 Hz with voicing beginning at 5 ms and an onset release burst during the first 10 ms. Although the stimulus does not contain a steady-state vowel, it is psychophysically perceived as a consonant–vowel speech syllable. By eliciting brainstem responses using a relatively shorter plosive-initial syllable /da/, BioMARK reflects the encoding of rapid acoustic offsets and onsets important for consonant identification.

Stimulus and recording parameters followed BioMARK protocol (Bio-logic, 2005), consistent with other studies (Banai et al., 2009; Dhar et al., 2009; Hornickel et al., 2009a; Krizman et al., 2010). As with the longer stimulus, responses were obtained to alternating polarity stimuli and then added together to extract the neural response from the cochlear microphonic and to eliminate stimulus artifact (Gorga et al., 1985). The /da/ was presented at a rate of 10.9 Hz, and two blocks of 3000 responses to each polarity were collected and averaged using a 74.67 ms time window (–15.8 ms pre-stimulus). Responses were sampled at 6857 Hz and band-pass filtered on-line from 100 to 2000 Hz, using a 12 dB/octave filter roll-off. Trials with artifact exceeding  $\pm 25 \mu\text{V}$  were excluded from the average. The 40 ms /da/ was presented via a magnetically shielded Bio-logic Systems insert earphone placed in the right ear at 80.3 dB SPL.

All responses were differentially recorded from Cz (active) to right earlobe (reference), with forehead as ground. During testing, the participants watched a captioned video of their choice with the sound level set at <40 dB SPL to facilitate a passive yet wakeful state.

## 2.3. Neurophysiologic analysis procedures

Speech-evoked brainstem responses were examined on a variety of measures in both time and frequency domains. Test–retest stability of the physiological measures was defined as no change from Tests 1 and 2 using a repeated measures analysis of variance with test session as the within-subject factor. Because of the variability in the duration of time between Tests 1 and 2, the duration of days between the recording sessions was used as a covariate in the analysis. Results of post hoc pairwise *t*-tests were adjusted for multiple comparisons by using a Bonferroni-correction.

### 2.3.1. Time domain analyses

**Overall response magnitude:** Root-mean-square (RMS) amplitude and signal-to-noise (SNR) of the response assessed overall response magnitude.

**Inter-response and stimulus-to-response correlations:** The effects of background noise on the timing of the response were quantified by performing inter-response (i.e., quiet vs. noise) correlations. This analysis was performed over the region of the response which included both the onset peaks and the FFR (5–180 ms). To assess how well the FFR represented the periodicity of the vowel in which the frequency of the  $F_0$  was constant, stimulus-to-response correlations were performed on the FFR (170 ms /da/: 50–180 ms, 40 ms /da/: 11–40 ms). Because Pearson's *r*-values were not normally distributed, stimulus-to-response and inter-response correlation measures were converted to Fisher's *z*'-scores prior to

subsequent parametric statistical analyses. For more detail on auditory brainstem responses to complex sounds, see Skoe and Kraus (2010).

**Discrete peak measures:** Measures of both timing and magnitude were utilized to assess the discrete peaks. Using the software with which the responses were collected (Neuroscan for the 170 ms /da/ and Bio-Logic for the 40 ms /da/), two experienced peak pickers manually marked the peaks of waves at the onset and transition portion in the averaged response in order to measure their latencies and amplitudes. In cases where the two peak pickers were not in agreement regarding the latency of a particular peak, a third experienced peak picker was consulted. This task was accomplished by selecting the peak with the largest positive or negative amplitude within the estimated time window that each was expected to occur (e.g., peak V: 9.00–10.00 ms and trough A: 10.00–11.00 ms for speech-evoked ABRs collected with Neuroscan) if it was visibly above the noise floor provided by the pre-stimulus period. When the background noise was introduced with the syllable, response peaks were often obscured or absent in the waveform, similar to the speech-evoked ABR collected in background noise in children (Russo et al., 2005). These peaks were omitted from statistical analyses.

### 2.3.2. Frequency domain analyses

Fast Fourier transform was used to evaluate the content and strength of frequency encoding of the frequency following portion of the response (i.e.,  $F_0$  and its harmonics). In the analysis of the 170 ms /da/ response, average spectral amplitude within a 40 Hz wide bin centered around the  $F_0$  and its harmonics were separately obtained from the transition (20–60 ms) and steady-state (60–180 ms) regions of the response. Separate spectral analysis was performed because the encoding of the  $F_0$  in the region corresponding to the formant transition is weakened by rapid shifts in higher formants (Johnson et al., 2008a; Hornickel et al., 2009b) compared to the sustained region in which the strength of  $F_0$  encoding is reinforced by unwavering formants that are integer multiples of the  $F_0$ . Moreover, because the frequency of the  $F_0$  was constant over a relatively long periodic portion compared to the 40 ms /da/, the spectral encoding of the frequency components was more focused and was encapsulated by the 40 Hz bins. For the 40 ms /da/, strength of frequency encoding was defined as the average spectral amplitude within a 100 Hz wide bin centered around the  $F_0$  and the harmonics. The larger bin size captured the broader spread of frequency representation primarily due to the absence of a steady-state vowel and a linearly rising  $F_0$  (103–125 Hz) in the voiced formant transition.

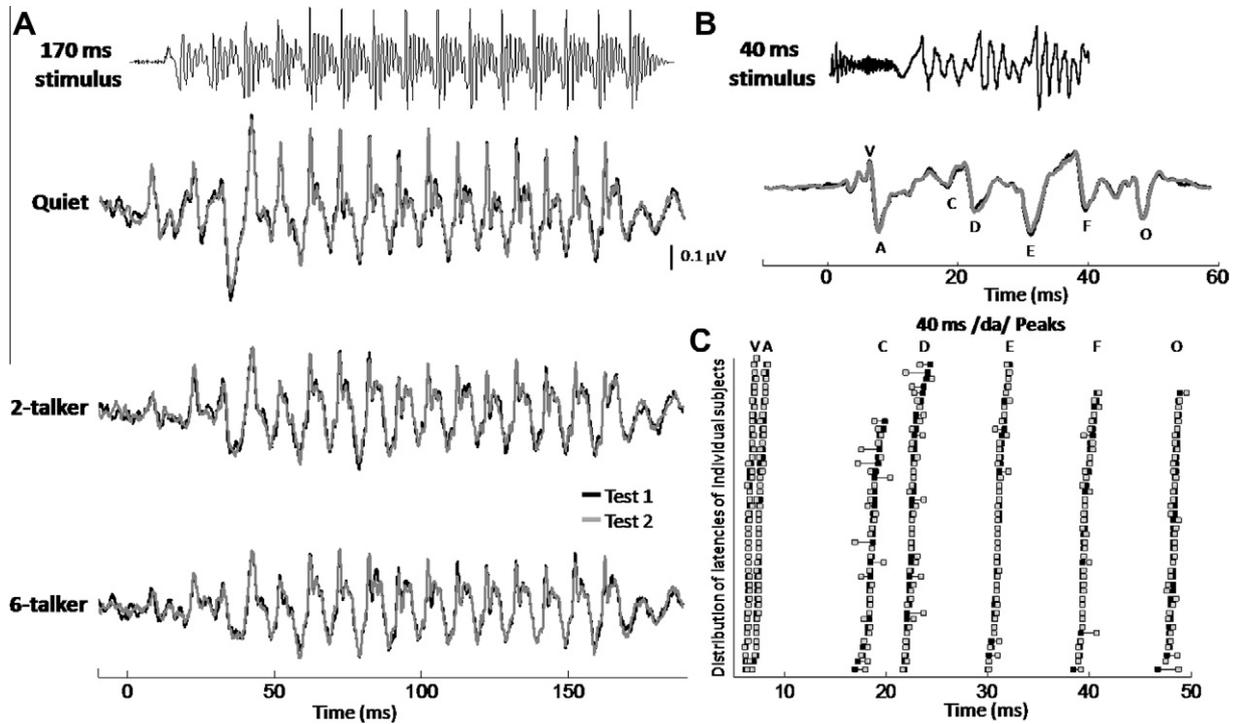
## 3. Results

### Study 1: Test–retest reliability of 170 ms /da/ in quiet and background noise

#### 3.1. Time domain analyses

##### 3.1.1. Overall response magnitude (SNR and RMS amplitude)

Grand average speech-evoked ABRs for both testing sessions in the three listening conditions: quiet, 2-talker and 6-talker babble are shown in Fig. 1A. Stability of the magnitude of the speech-evoked brainstem response was evaluated over the entire range of the response (0–180 ms). The global magnitude of neural activation was replicable from test to retest sessions for each subject and was consistently vulnerable to significant degradation in background noise (see Table 1 for means and standard deviations). A two session (Test 1 vs. Test 2)  $\times$  three condition (quiet, 2-talker and 6-talker babble) two-way repeated measures ANOVA was



**Fig. 1.** (A) Overlay of grand average brainstem responses from Test 1 (black) and Test 2 (grey) to the 170 ms /da/ stimulus without background noise (top) and in two background noise conditions, 2-talker (middle) and 6-talker (bottom) babble. (B) Acoustic waveform of the 40 ms /da/ (top). Overlay of grand average brainstem responses from Test 1 (black) and Test 2 (grey) to the 40 ms /da/ stimulus. Transient peaks (V, A, C and O) and steady-state peaks (D, E and F) are labeled in the response. (C) Dot plot of the individual subjects' Test 1 (black square) and Test 2 (grey square) latencies of each peak in the response (V, A, C, D, E, F and O) with a horizontal line connect the Tests 1 and 2 values. Here, and in all figures showing individual-subject data, subjects are sorted by pre-test values independently in each plot.

performed separately on each of the two dependent variables which reflected overall response magnitude (i.e., SNR and RMS amplitude). There were no significant main effects of SNR ( $F = 1.726, p = 0.199$ ), RMS amplitude ( $F = 0.180, p = 0.674$ ) nor interaction between test session and condition (SNR:  $F = 0.095, p = 0.910$ , RMS:  $F = 0.131, p = 0.877$ ). However there was a significant main effect of condition (SNR:  $F = 18.636, p < 0.0001$ ; RMS:  $F = 10.441, p < 0.0001$ ). Post hoc pairwise *t*-tests showed that 2- and 6-talker background noise significantly degraded both measures of overall magnitude compared to quiet during Test 1 (RMS:  $t = 3.145, 4.437, t < 0.0001$ ; SNR:  $t = 3.660, 3.981, p = 0.001, < 0.0001$ , respectively) and Test 2 (RMS:  $t = 2.518, 3.735, p = 0.017, 0.001$ ; SNR:  $t = 2.993, 4.041, p = 0.005, < 0.0001$ ).

**3.1.2. Quiet-to-noise inter-response analysis (5–180 ms)**

The relationships among brainstem responses recorded in quiet and background noise were explored using cross-correlation analysis. The impact of noise on the speech-evoked brainstem response for each subject was similar across test sessions. The stability of the

degree of similarity between the response across listening conditions was analyzed performing a two-way repeated measures ANOVA separately for each of the two dependent variables arising from quiet-to-noise correlation analysis (i.e., lag time and *z*'-score of Pearson *r*-value). A two session (Test 1 vs. Test 2) × two condition (quiet vs. 2-talker and quiet vs. 6-talker babble) repeated measures ANOVA for lag time showed no significant main effects of session or condition ( $F = 0.300, 0.385, p = 0.588, 0.540$ , respectively) nor a significant interaction between session and condition ( $F = 0.385, p = 0.540$ ). Moreover, a two-way repeated measures ANOVA of the *z*'-score of the Pearson's *r*-value showed no significant main effect of session ( $F = 1.734, p < 0.198$ ) nor a significant interaction between session and condition ( $F = 0.243, p = 0.626$ ). However, there was a significant main effect of condition ( $F = 19.627, p < 0.0001$ ), with post hoc pairwise *t*-tests showing that the 6-talker babble consistently imposed significantly greater challenge than the 2-talker noise in the neural phase-locking of periodic information found in the stimulus as shown by a lower Pearson's *r*-value within a given test session (Test 1:  $t = 3.532, p = 0.001$ , Test 2:  $t = 3.472, p = 0.002$ ). See Table 2 and Fig. 2 for means and standard deviations as well as individual test-rest data.

**Table 1**

Mean ( $\pm 1$  SD) RMS and SNR of the 170 ms /da/ response for each listening condition (Quiet, 2- and 6-talker babble) from Tests 1 and 2.

		170 ms /da/	
		Overall response magnitude	
		Test 1	Test 2
		Mean $\pm$ SD	Mean $\pm$ SD
RMS	Quiet	0.14 $\pm$ 0.04	0.14 $\pm$ 0.03
	2-Talker	0.13 $\pm$ 0.04	0.13 $\pm$ 0.03
	6-Talker	0.13 $\pm$ 0.03	0.13 $\pm$ 0.03
SNR	Quiet	1.86 $\pm$ 0.33	1.90 $\pm$ 0.45
	2-Talker	1.65 $\pm$ 0.40	1.68 $\pm$ 0.40
	6-Talker	1.56 $\pm$ 0.41	1.63 $\pm$ 0.34

**3.1.3. Stimulus-to-response correlation (50–180 ms)**

The stability of the degree of similarity between the FFR and the stimulus and the time displacement that produced the highest Pearson's *r*-value were assessed by using stimulus-to-response correlations. The fidelity of the FFR in encoding the timing features of the stimulus was replicable and demonstrated good intrasubject test-retest reliability. Moreover, these properties were consistently affected by noise over multiple test sessions; see Table 3 and Fig. 3 for means and standard deviations as well as individual test-retest data. To determine the reproducibility of response fidelity to the stimulus from one session to another, a two session (Test 1 vs. Test 2) × three conditions (quiet, 2- and 6-talker babble)

**Table 2**

Mean ( $\pm$  SD) lag time and Pearson's  $r$ -value derived from an inter-response analysis (i.e., Quiet vs. 2-Talker and Quiet vs. 6-Talker) from Tests 1 and 2.

		Inter-response measures	
		Test 1 Mean $\pm$ SD	Test 2 Mean $\pm$ SD
Lag time (ms)	Quiet vs. 2-Talker	0.26 $\pm$ 0.16	0.25 $\pm$ 0.16
	Quiet vs. 6-Talker	0.28 $\pm$ 0.27	0.27 $\pm$ 0.20
Pearson's $r$ -value	Quiet vs. 2-Talker	0.66 $\pm$ 0.15	0.67 $\pm$ 0.14
	Quiet vs. 6-Talker	0.62 $\pm$ 0.15	0.63 $\pm$ 0.14

**Table 3**

Mean ( $\pm$  SD) lag time and Pearson's  $r$ -value derived from stimulus-to-response analysis for each listening condition from Tests 1 and 2.

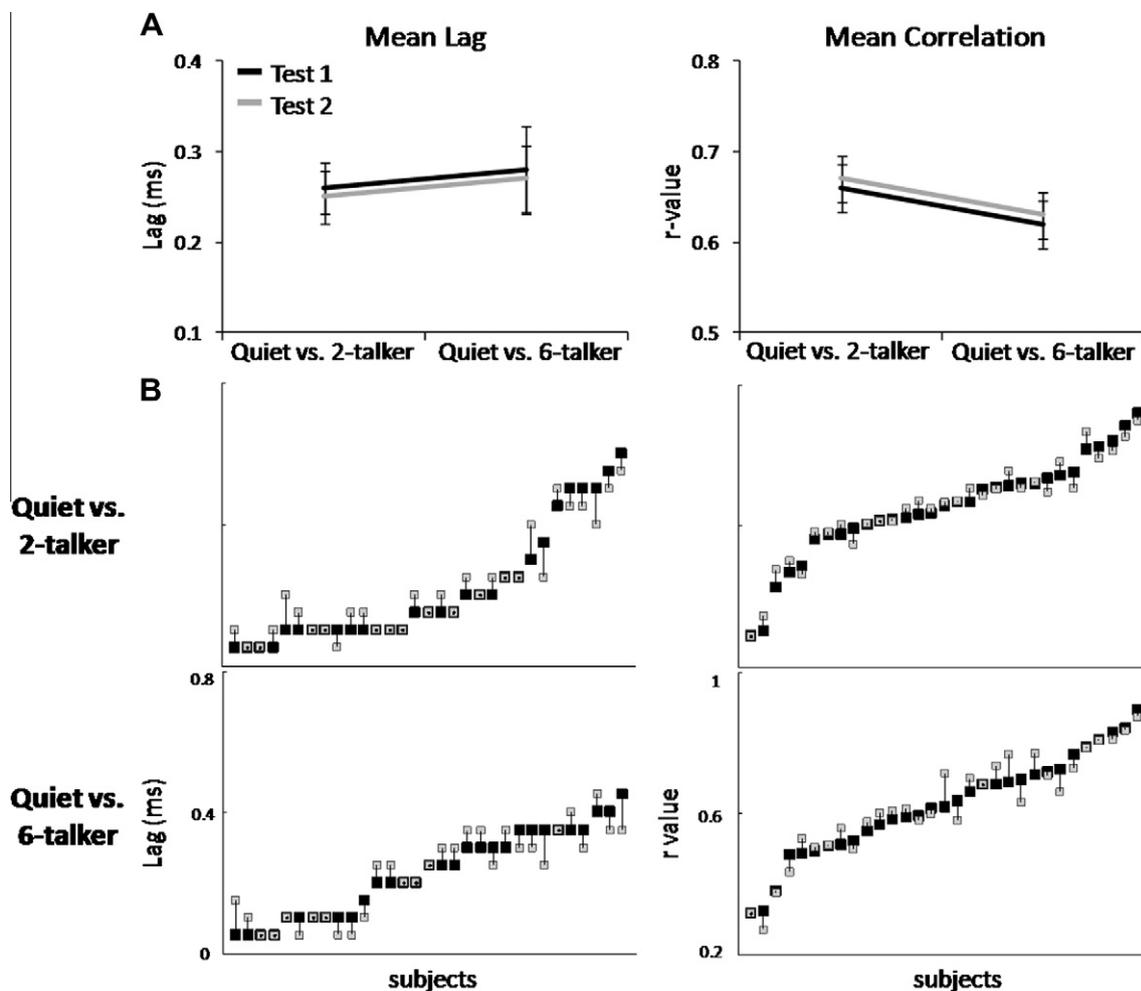
		Stimulus-to-response measures	
		Test 1 Mean $\pm$ SD	Test 2 Mean $\pm$ SD
Lag time (ms)	Quiet	8.91 $\pm$ 0.58	8.93 $\pm$ 0.62
	2-Talker	9.19 $\pm$ 0.75	9.19 $\pm$ 0.74
	6-Talker	9.25 $\pm$ 0.83	9.21 $\pm$ 0.78
Pearson's $r$ -value	Quiet	0.26 $\pm$ 0.07	0.26 $\pm$ 0.07
	2-Talker	0.21 $\pm$ 0.07	0.21 $\pm$ 0.07
	6-Talker	0.22 $\pm$ 0.07	0.21 $\pm$ 0.06

two-way repeated measures ANOVA was performed separately for each of the two dependent variables of the stimulus-to-response analysis (lag time and  $z'$ -score of Pearson  $r$ -value). There was no significant main effect of session for either the lag time or  $z'$ -score measures ( $F = 0.011, 1.529, p = 0.915, 0.226$ , respectively) nor a significant interaction between session and condition (lag time:  $F = 1.208, p = 0.313, z'$ -score:  $F = 0.127, p = 0.881$ ). However, there was a main effect of condition for both measures (lag time:  $F = 7.045, p < 0.003, z'$ -score:  $F = 17.082, p < 0.0001$ ). Post hoc pairwise  $t$ -tests showed that while the stimulus-to-response measures of lag time and  $z'$ -score did not differ from test to retest in any of the three listening conditions (lag time:  $t = -0.949, -0.309, 1.062, p = 0.350, 0.759, 0.297$ ;  $z'$ -score:  $t = -1.102, -0.490, -1.022, p = 0.279, 0.628, 0.315$ , for quiet, 2-talker, and 6-talker, respectively), the presence of 2- and 6-talker background noise sig-

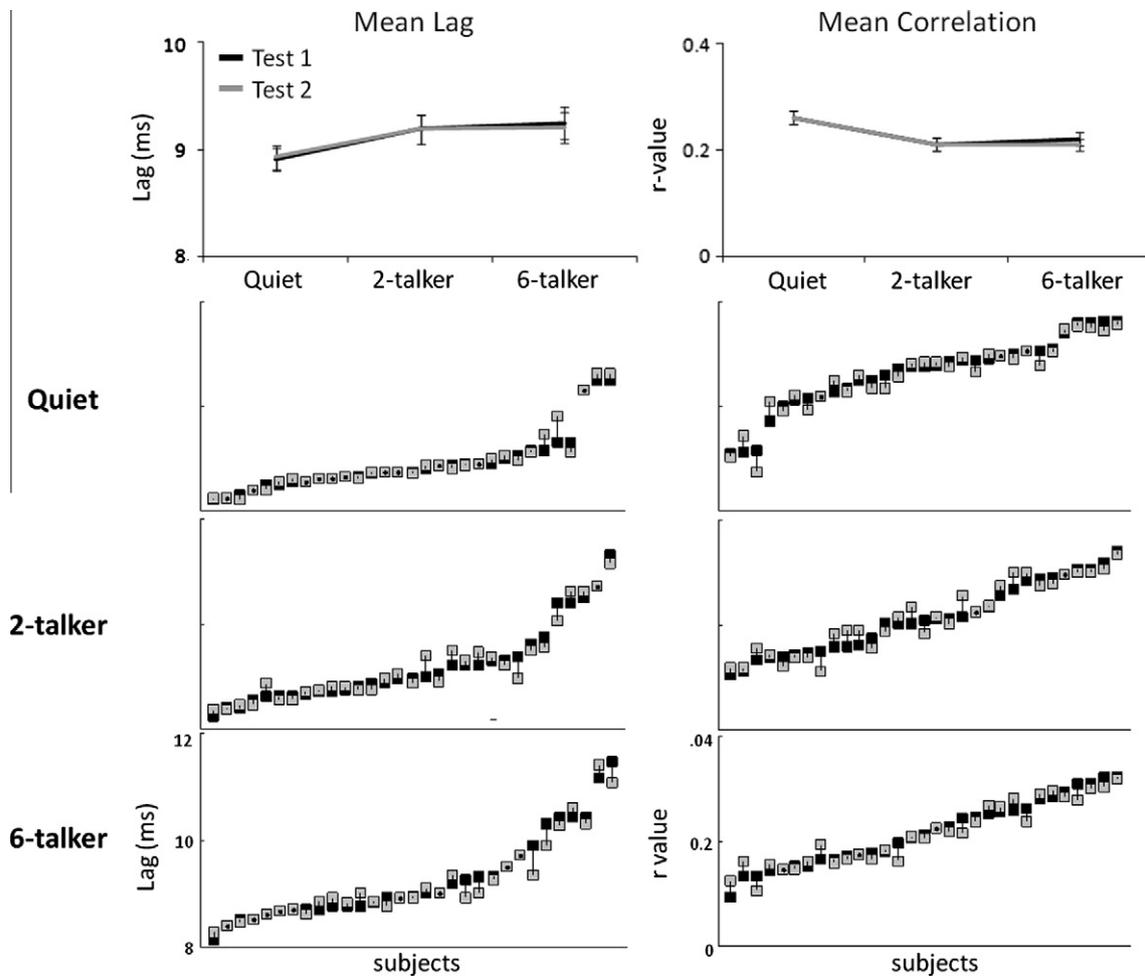
nificantly degraded both measures compared to quiet during both Test 1 (lag time:  $t = -3.499, 3.597, p = 0.001, 0.001$ ;  $z'$ -score:  $t = -5.682, -5.306, p < 0.0001, < 0.0001$ ) and Test 2 (lag time:  $t = -3.752, -3.584, p = 0.001, 0.001$ ;  $z'$ -score:  $t = -5.273, -4.881, p < 0.0001, < 0.0001$ ). However, there were no significant differences between 2- and 6-talker babble for either lag time or  $z'$ -score calculated from the responses within each test session (Test 1: lag time:  $t = -1.472, p = 0.151, z'$ -score:  $t = 0.916, p = 0.367$ ; Test 2: lag time:  $t = -0.450, p = 0.656, z'$ -score:  $t = 0.095, p = 0.925$ ).

### 3.1.4. Discrete peak measures

**Onset peaks (Waves V and A): Detectability:** In the quiet condition, the onset response peaks (Waves V and A) were detected in all 31 participants' responses in Tests 1 and 2. In the 2-talker



**Fig. 2.** (A) Mean ( $\pm$  SEM) lag time and correlation from Test 1 (black) and Test 2 (grey) derived from inter-response comparisons for the 170-ms /da/ (Quiet vs. 2-talker and Quiet vs. 6-talker). (B) Individual subjects' Test 1 (black square) and Test 2 (grey square) lag times and  $r$ -values with vertical lines connecting Tests 1 and 2 values.



**Fig. 3.** (A) Mean ( $\pm 1$  SEM) lag and correlation from Test 1 (black) and Test 2 (grey) derived from the stimulus-to-response comparisons for the 170-ms /da/ (Quiet, 2-talker, 6-talker). (B) Individual subjects' Test 1 (black square) and Test 2 (grey square) lag times and *r*-values with vertical lines connecting Tests 1 and 2 values.

condition, Waves V and A were present in 17 participants' responses during Tests 1 and 2. For one participant, the onset response present in Test 1 was absent in Test 2 and conversely, for another participant, it was present in Test 2 but absent in Test 1. In the 6-talker condition, the onset response was present in 15 participants' responses during Tests 1 and 14 during Test 2. The responses of two participants who exhibited an onset response during Test 1 did not show it during Test 2. One participant showed an onset response during Test 2 which was absent during Test 1. Pearson's  $\chi^2$  showed that detectability of the onset waves was not statistically different from Tests 1 and 2 in either the 2-talker ( $p = 0.108$ ) or 6-talker ( $p = 0.252$ ) conditions.

Response latency: A two session (Test 1 vs. Test 2)  $\times$  three condition (quiet, 2-talker and 6-talker) two-way repeated measures

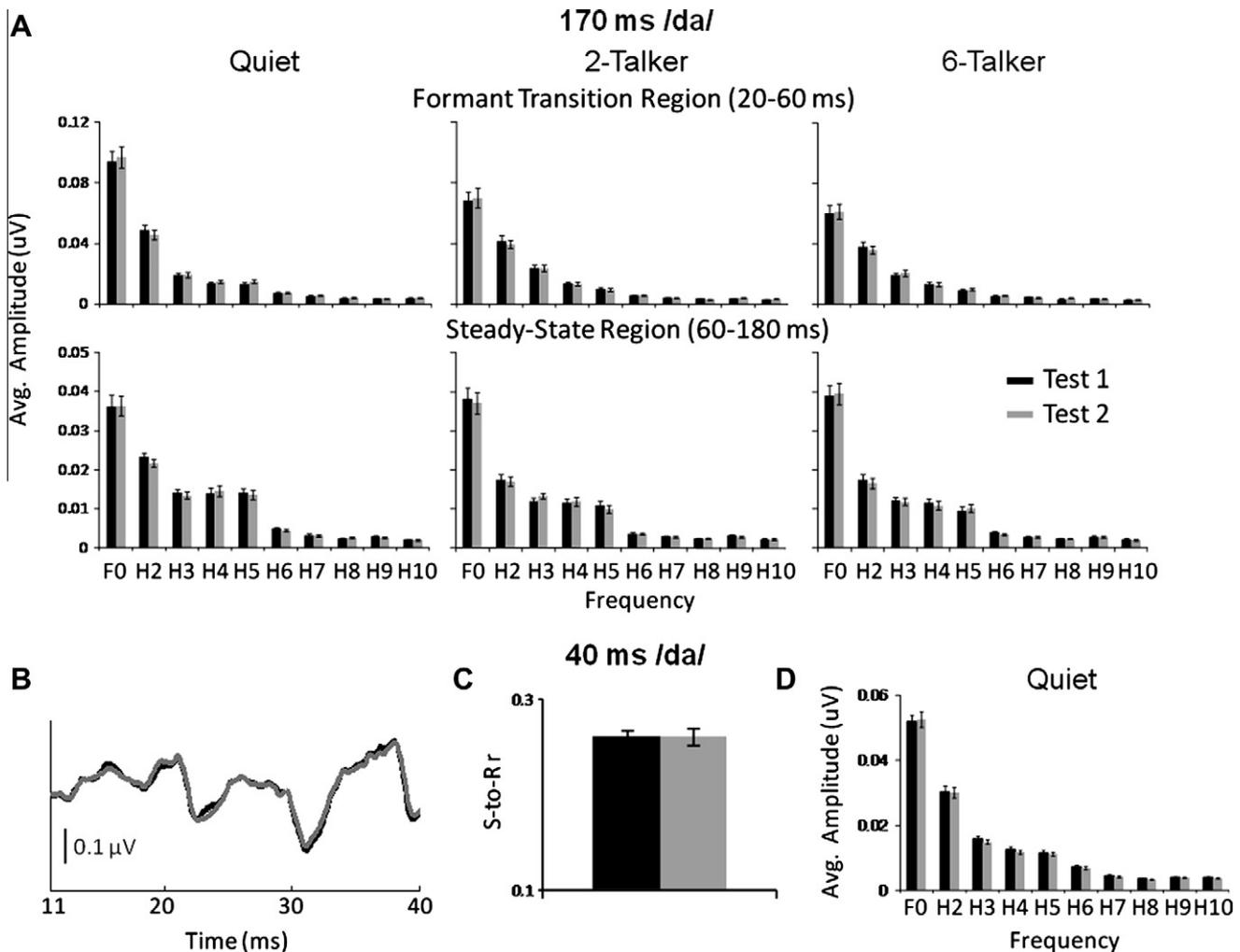
MANOVA was performed with Waves V and A latencies as the dependent variables. There was no significant main effect of session ( $F = 2.109, p = 1.84$ ) nor significant interaction between session and condition ( $F = 3.257, p = 0.096$ ). As expected, there was a main effect of condition ( $F = 19.229, p = 0.001$ ) suggesting that the desynchronizing influence of noise on neural firing resulted in a later latency, see Table 4 for means and standard deviations of Wave V and A latencies. In summary, neither the intrasubject detectability nor the latencies of V or A differed significantly from test to retest.

*Formant transition period (20–60 ms):* Peaks in the transition region of the brainstem response collected in either quiet or background noise, were replicable over a test–retest period. This result is important as the neural timing of these peaks, particularly in background noise, has been found to be associated with

**Table 4**

Mean ( $\pm 1$  SD) latencies of peaks (positive and negative) in the onset and transition regions for each listening condition from Tests 1 and 2.

		Quiet		2-Talker babble		6-Talker babble	
		Test 1 Mean $\pm$ SD	Test 2 Mean $\pm$ SD	Test 1 Mean $\pm$ SD	Test 2 Mean $\pm$ SD	Test 1 Mean $\pm$ SD	Test 2 Mean $\pm$ SD
Positive peak Latency (ms)	V	9.57 $\pm$ 0.58	9.56 $\pm$ 0.57	9.99 $\pm$ 0.43	9.99 $\pm$ 0.45	10.23 $\pm$ 0.48	10.23 $\pm$ 0.50
	33	33.21 $\pm$ 0.46	33.19 $\pm$ 0.51	33.81 $\pm$ 0.59	33.82 $\pm$ 0.62	34.34 $\pm$ 0.64	34.34 $\pm$ 0.67
	43	43.31 $\pm$ 0.46	43.30 $\pm$ 0.47	43.33 $\pm$ 0.56	43.33 $\pm$ 0.60	43.45 $\pm$ 0.75	43.49 $\pm$ 0.74
	53	53.48 $\pm$ 0.68	53.47 $\pm$ 0.71	53.46 $\pm$ 0.79	53.47 $\pm$ 0.76	53.56 $\pm$ 0.83	53.60 $\pm$ 0.77
Negative peak Latency (ms)	A	10.92 $\pm$ 0.88	10.93 $\pm$ 0.89	11.29 $\pm$ 0.60	11.29 $\pm$ 0.59	11.31 $\pm$ 0.64	11.32 $\pm$ 0.68
	35	35.28 $\pm$ 0.64	35.29 $\pm$ 0.58	35.58 $\pm$ 0.58	35.60 $\pm$ 0.62	35.81 $\pm$ 0.63	35.83 $\pm$ 0.64
	45	45.55 $\pm$ 0.63	45.54 $\pm$ 0.64	45.53 $\pm$ 0.59	45.54 $\pm$ 0.63	45.51 $\pm$ 0.75	45.29 $\pm$ 0.78
	56	56.09 $\pm$ 0.59	56.01 $\pm$ 0.61	56.20 $\pm$ 0.55	56.23 $\pm$ 0.73	56.22 $\pm$ 0.77	56.22 $\pm$ 0.76



**Fig. 4.** (A) Average fundamental frequency ( $F_0$ ) and harmonic (H2–H10) amplitudes ( $\pm 1$  SEM) of the transition region (20–60 ms, top) and steady-state region (60–180 ms, bottom) of the response elicited by the 170 ms /da/ for each listening condition. (B) Overlay of grand average responses from the FFR region (11–40 ms) from Test 1 (black) and Test 2 (grey) for the 40 ms /da/. (C) Mean stimulus-to-response correlation values ( $\pm 1$  SEM) for the 40 ms /da/ from Tests 1 and 2. (D) Mean fundamental frequency ( $F_0$ ) and harmonic (H2–H10) amplitudes ( $\pm 1$  SEM) of the frequency following region (11–40 ms) of the response elicited by the 40 ms /da/.

the speech perceptual ability (Anderson et al., 2010). Table 4 lists the means and standard deviations of peak latencies from the formant transition period. Analysis was conducted on the latencies of peaks in the formant transition period (three positive- and negative-going peak pairs at mean latencies of approximately 33, 35, 43, 45, 53 and 56 ms). The voicing onset peaks, occurring approximately at 23 and 24 ms after the response onset, were excluded from analysis as their amplitudes did not exceed the noise floor, especially in the babble conditions. A two session (Test 1 vs. Test 2)  $\times$  three condition (quiet, 2-talker and 6-talker babble) two-way repeated measures MANOVA was performed with the latencies of these peaks as dependent variables. Results showed no significant main effects of session ( $F = 0.740$ ,  $p = 0.642$ ) or condition ( $F = 1.817$ ,  $p = 0.117$ , respectively) nor a significant interaction between session and condition ( $F = 1.635$ ,  $p = 0.161$ ).

### 3.2. Frequency domain analysis

#### 3.2.1. Representation of fundamental and formant frequencies

The strength of frequency encoding in the transition and steady-state regions of the response elicited by the different listening conditions was examined in the frequency domain using the fast

Fourier transform (FFT). Individual responses were segmented into two time ranges: (1) 20–60 ms, which includes the response to the formant transition of the stimulus and (2) 60–180 ms, which includes the response to the steady-state vowel. To examine the strength of frequency encoding, average response magnitudes were calculated for 40 Hz wide bins surrounding the  $F_0$  (100 Hz) and subsequent nine harmonics (200 Hz, 300 Hz ... 1000 Hz). These test-retest results indicate that the strength of frequency encoding was stable and was not susceptible to change for each subject; refer to Fig. 4A for the average and standard deviation of FFT amplitudes from Tests 1 and 2 for each listening condition and response time range. A two session (Test 1 vs. Test 2)  $\times$  three condition (quiet, 2-talker and 6-talker babble) two-way repeated measures ANOVA was separately performed on the formant transition (20–60 ms) and steady-state (60–180 ms) regions of the response. There was no significant main effect of test session ( $F = 1.168$ ,  $1.585$ ,  $p = 0.366$ ,  $0.179$ , respectively) nor a significant interaction between test and condition ( $F = 0.538$ ,  $0.725$ ,  $p = 0.886$ ,  $0.744$ , respectively). As expected, there was a significant main effect of condition ( $F = 3.564$ ,  $3.909$ ,  $p = 0.022$ ,  $0.012$ ), indicating that the presence of multi-speaker babble degraded the subcortical encoding of frequencies compared to responses recorded in quiet.

**Study 2: Test–retest reliability of the 40 ms /da/**

**3.3. Time domain analysis**

**3.3.1. Overall response magnitude (SNR and RMS amplitude)**

Grand average responses are shown in Fig. 1B. Stability of the overall magnitude of the response was evaluated over the entire range of the response from test to retest. The global magnitude of neural activation is replicable and is resistant to gain or reduction over time for each subject, see Table 5 for means and standard deviations. A two session (Test 1 vs. Test 2) one-way repeated measures ANOVA was performed on each of the two dependent variables measuring overall response magnitude (i.e., SNR and RMS amplitude). There was no significant main effect of session for SNR ( $F = 0.107, p = 0.745$ ) or RMS amplitude ( $F = 0.489, p = 0.488$ ).

**3.3.2. Stimulus-to-response analysis**

The FFR (11–40 ms) was evident in all subjects, see Fig. 4B for an overlay of grand average responses in this time region from both testing sessions. The stability of the degree of similarity between the FFR and the stimulus as well the time displacement that produced the highest  $r$ -value was assessed using stimulus-to-response correlations. Intrasubject response fidelity to the stimulus was replicable and demonstrated good test–retest reliability, see Fig. 4C for means and standard deviations. A two session (Test 1 vs. Test 2) one-way repeated measures ANOVA was performed on each of the two dependent variables which measured the precision to which the FFR mimics the stimulus (i.e., lag time and  $z$ -score of the Pearson’s  $r$ -value). There was no significant main effect of session for lag time ( $F = 1.615, p = 0.211$ ) or the  $z$ -score of the Pearson’s  $r$ -value ( $F = 0.006, p = 0.941$ ).

**3.3.3. Peak analysis**

Latencies and amplitudes of peaks V, A, C, D, E, F and O, as well as composite VA measures (i.e., duration, amplitude and slope and area) obtained during Test 1 were comparable with those obtained during Test 2; within-subject latency values are shown in Fig. 1C. Peak responses fall into two categories: those that encode transient events in the stimulus (V, A, C and O), and those that encode the periodicity of the vowel (D, E and F) (Russo et al., 2004; Kraus and Nicol, 2005; Dhar et al., 2009; Hornickel et al., 2009a; Krizman

**Table 5**  
Mean ( $\pm 1$  SD) RMS and SNR of the 40 ms /da/ response from Tests 1 and 2.

	40 ms /da/	
	Overall response magnitude	
	Test 1 Mean $\pm$ SD	Test 2 Mean $\pm$ SD
RMS	0.09 $\pm$ 0.02	0.09 $\pm$ 0.02
SNR	3.01 $\pm$ 1.11	2.89 $\pm$ 1.22

**Table 6**  
Mean ( $\pm 1$  SD) latencies and amplitudes of the 40 ms /da/ response peaks (V, A, C, D, E, F and O) from Tests 1 and 2.

Peak	Latency		Amplitude	
	Test 1	Test 2	Test 1	Test 2
	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD
V	6.65 $\pm$ 0.27	6.68 $\pm$ 0.27	0.13 $\pm$ 0.05	0.13 $\pm$ 0.04
A	7.62 $\pm$ 0.35	7.62 $\pm$ 0.37	-0.20 $\pm$ 0.06	-0.21 $\pm$ 0.06
C	18.60 $\pm$ 0.68	18.47 $\pm$ 0.68	-0.03 $\pm$ 0.06	-0.03 $\pm$ 0.05
D	22.67 $\pm$ 0.59	22.72 $\pm$ 0.58	-0.13 $\pm$ 0.07	-0.14 $\pm$ 0.07
E	31.12 $\pm$ 0.53	31.20 $\pm$ 0.57	-0.22 $\pm$ 0.06	-0.21 $\pm$ 0.07
F	39.70 $\pm$ 0.57	39.71 $\pm$ 0.50	-0.14 $\pm$ 0.09	-0.13 $\pm$ 0.08
O	48.26 $\pm$ 0.43	48.34 $\pm$ 0.39	-0.15 $\pm$ 0.06	-0.16 $\pm$ 0.06

**Table 7**  
Mean ( $\pm 1$  SD) VA measures (i.e., duration, amplitude, slope and area) of the 40 ms /da/ onset response from Tests 1 and 2.

	VA measures	
	Test 1	Test 2
	Mean $\pm$ SD	Mean $\pm$ SD
Duration	0.98 $\pm$ 0.23	0.94 $\pm$ 0.21
Amplitude	0.33 $\pm$ 0.09	0.34 $\pm$ 0.08
Slope	-0.35 $\pm$ 0.11	-0.37 $\pm$ 0.12
Area	0.16 $\pm$ 0.05	0.15 $\pm$ 0.05

et al., 2010). For analysis, the peaks were divided in this way. The timing and robustness of these response peaks were highly stable and were reliably replicated over time for each subject. First, a one-way 2 session (Test 1 vs. Test 2) repeated measures MANOVA was performed with V, A, C and O latency and amplitude and VA duration, amplitude, slope and area serving as dependent variables. There were no significant main effects of session ( $F = 0.989, p = 0.500$ ). A second one-way two session (Tests 1 vs. 2) repeated measures ANOVA was performed with D, E and F latency and amplitude serving as dependent variables. There were no significant main effects of session ( $F = 1.436, p = 0.231$ ), refer to Tables 6 and 7 for means and standard deviations.

**3.4. Frequency domain analysis**

**3.4.1. Representation of fundamental and formant frequencies**

The strength of frequency encoding in the FFR (11–40 ms) elicited by the 40 ms /da/ obtained during Test 1 was comparable to that obtained during Test 2, see Fig. 4D. A two session (Test 1 vs. Test 2) one-way repeated measures MANOVA with the average response magnitudes from each of the 10 frequencies bins reflecting the F0 (100 Hz) and the subsequent harmonics (200, 300 . . . 1 kHz, with a 100 Hz bin size) serving as dependent variables of frequency encoding was performed to examine the reliability of frequency encoding over the two separate recording sessions. There were no significant main effect of session ( $F = 0.560, p = 0.834$ ). Thus, magnitude of frequency encoding of the periodic elements of this stimulus was stable for each subject over time.

**4. Discussion**

Overall, we found high test–retest reliability of the brainstem response to speech syllables recorded in quiet and noisy listening conditions in young adults. There were no significant test–retest differences in measures that quantify the response in either time nor frequency domains regardless of the type of stimulus used (40 ms vs. 170 ms /da/) or background listening conditions (quiet vs. 2- and 6-talker babble). Demonstrating stability of the response is critical, especially as the speech-evoked brainstem response plays an increasingly important role in the objective assessment of auditory processing.

Like the click-evoked brainstem response, the speech-evoked brainstem response also possesses the attribute of having a highly predictable nature shown by its remarkable degree of both temporal precision and test–retest reliability in young adults. For instance, the stop-consonant of both the 40 ms and 170 ms /da/, characterized to be brief and spectrally stochastic, consistently elicited an onset response in the quiet condition at the expected latencies for all subjects. This finding points to the reliability of these speech stimuli to elicit an onset response as well as the robust nature of the neural encoding of young adults to consistently encode this feature from test to retest. Also, for both stimuli, the region of the response reflecting the encoding of the formant transition in the stimulus (analogous regions between the short

and long /da/ stimuli), showed faithful representation of stimulus timing corresponding to  $F_0$  and formants during both test sessions. This attribute allows for accurate and valid evaluation of the speech-evoked brainstem activity elicited in quiet or noisy backgrounds over multiple test sessions. Thus, due to the steadiness of the response, changes that are observed between the initial test and retest, in a study involving auditory training, can be interpreted in terms of training related outcomes. Additionally, tracking changes in a physiologic response such as the speech-evoked ABR can be used to objectively evaluate the effectiveness of a particular training program (Tzounopoulos and Kraus, 2009; Kraus and Chandrasekaran, 2010), thereby enhancing the clinical utility of aural rehabilitation programs.

The high test–retest reliability of the speech-evoked ABR obtained in the present study is commensurate with a previous study conducted in children. In a study conducted on eight normal hearing children whose speech-evoked brainstem responses were obtained on two separate sessions spaced 2–10 months apart, Russo and colleagues (2004) found that most brainstem measures did not change significantly over the test–retest time interval. Exceptions included the VA interpeak amplitude and slope in quiet. While in the present study, these measures did not differ at retest, these exceptions may be partly explained by known factors, such as the variability of the onset response amplitude (Starr and Don, 1988), as well as the developmental changes that are observed in children for the auditory brainstem response to sounds composed of acoustic elements relevant to speech, often manifested as delayed and less synchronous onset responses (Johnson et al., 2008b).

We have shown that the test–retest reliability of the speech-evoked ABR test was excellent across listening conditions and stimulus durations, with no significant differences at the group level and high levels of agreement at the individual level. Given the high test–retest reliability, the speech-evoked ABR test holds promise as a useful investigative instrument to quantify with confidence the degree to which the neural system of young adults is consistent in encoding a complex sound at the preconscious level. Establishing this level of stability has positive broad implications on research and clinical assessment whenever auditory processing is of interest. This includes investigations of ABRs to complex sounds in challenging listening conditions in populations with auditory specialization (e.g., musicians, native language speakers) and the management of auditory deficits (e.g., auditory processing disorders, language-based learning impairments, hearing loss and age-related hearing decline).

## Acknowledgements

The authors wish to thank Professor Ann Bradlow for the use of her background noise stimuli and Professor Steven Zecker for his advice on the statistical treatment of these data. We would also like to thank the people who participated in this study. This work was supported by RO1 DC01510, T32 NS047987 and F32 DC008052.

## References

Akhoun I, Gallégo S, Moulin A, Menard M, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol* 2008;119:922–33.

Anderson S, Chandrasekaran B, Skoe E, Kraus N. Neural timing is linked to speech perception in noise. *J Neurosci* 2010;30:4922–6.

Banai K, Hornickel JM, Skoe E, Nicol T, Zecker S, Kraus N. Reading and subcortical auditory function. *Cereb Cortex* 2009;19:2699–707.

Bidelman G, Gandour JT, Krishnan A. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J Cogn Neurosci* 2009;175:196–205.

Bio-logic SC. Auditory Evoked Potential (AEP) System User's and Service Manual. 590-AEPUM1, Rev C; 2005.

Boston JR, Møller AR. Brainstem auditory-evoked potentials. *Crit Rev Biomed Eng* 1985;13:97–123.

Brandt J, Rosen JJ. Auditory phonemic perception in dyslexia – categorical identification and discrimination of stop consonants. *Brain Lang* 1980;9:324–37.

Brown L, Sherbenou RJ, Johnsen SK. Test of nonverbal intelligence: a language-free measure of cognitive ability. Austin, TX: PRO-ED Inc.; 1997.

Chandrasekaran B, Kraus N. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 2010;47:236–46.

Cone-Wesson B, Kurtzberg D, Vaughan HG. Electrophysiologic assessment of auditory pathways in high-risk infants. *Int J Pediatr Otorhinolaryngol* 1987;14:203–14.

Cone-Wesson B, Vohr BR, Sininger YS, Widen JE, Folsom RC, Gorga MP, Norton SJ. Identification of neonatal hearing impairment: infants with hearing loss. *Ear Hear* 2000;21:488–507.

Cunningham J, Nicol T, King C, Zecker SG, Kraus N. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hear Res* 2002;169:97–111.

Dhar S, Abel R, Hornickel J, Nicol T, Skoe E, Zhao W, Kraus N. Exploring the relationship between physiological measures of cochlear and brainstem function. *Clin Neurophysiol* 2009;120:959–66.

Edwards RM, Buchwald JS, Tanguay PE, Schwafel JA. Sources of variability in auditory brain-stem evoked-potential measures over time. *Electroencephalogr Clin Neurophysiol* 1982;53:125–32.

Galbraith GC, Amaya EM, Diaz de Rivera JM, Donan NM, Duong MT, Hsu JN, Tran K, Tsang LP. Brain stem evoked response to forward and reversed speech in humans. *NeuroReport* 2004;15:2057–60.

Galbraith GC, Arbagey PW, Branski R, Comerici N, Rector PM. Intelligible speech encoded in the human brain stem frequency following response. *NeuroReport* 1995;6:2363–7.

Galbraith GC, Jhaveri SP, Kuo J. Speech-evoked brainstem frequency-following responses during verbal transformations due to world repetition. *Electroencephalogr Clin Neurophysiol* 1997;102:46–53.

Gorga M, Abbas P, Worthington D. Stimulus calibration in ABR measurements. In: Jacobsen J, editor. *The auditory brainstem response*. San Diego: College Hill Press; 1985. p. 49–62.

Hall JW. *Handbook of auditory evoked responses*. Heedham Heights, MA: Allyn and Bacon; 1992.

Hood LJ. *Clinical applications of the auditory brainstem response*. San Diego, CA: Singular Publishing Group, Inc.; 1998.

Hornickel J, Skoe E, Kraus N. Subcortical laterality of speech encoding. *Audiol Neurotol* 2009a;14:198–207.

Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Subcortical differentiation of voiced stop consonants: relationships to reading and speech in noise perception. *Proc Natl Acad Sci* 2009b;106:13022–7.

Jacobson J. *The auditory brainstem response*. San Diego: College Hill Press; 1985.

Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E, Kraus N. Brainstem encoding of voiced consonant–vowel stop syllables. *Clin Neurophysiol* 2008a;119:2623–35.

Johnson KL, Nicol T, Zecker SG, Kraus N. Developmental plasticity in the human auditory brainstem. *J Neurosci* 2008b;28:4000–7.

Klatt D. Software for cascade/parallel formant synthesizer. *J Acoust Soc Am* 1980;67:971–5.

Kraus N, Chandrasekaran B. Musical training for the development of auditory skills. *Nat Neurosci* 2010;11:599–605.

Kraus N, Nicol T. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci* 2005;28:176–81.

Krishnan A, Swaminathan J, Gandour JT. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci* 2009;21:1092–105.

Krishnan A, Xu YS, Gandour JT, Cariani PA. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 2004;189:1–12.

Krizman J, Skoe E, Kraus N. Stimulus rate and subcortical auditory processing of speech. *Audiol Neurotol* 2010;15:332–42.

Lee K, Skoe E, Kraus N, Ashley R. Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 2009;29:5832–40.

Møller AR, Jannetta P. Neural generators of the auditory brainstem response. In: Jacobsen J, editor. *The auditory brainstem response*. San Diego, CA: College Hill Press; 1985. p. 13–32.

Moushegian G, Rupert AL, Stillman RD. Scalp-recorded early responses in man to frequencies in speech range. *Electroencephalogr Clin Neurophysiol* 1973;35:665–7.

Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci* 2007;104:15894–8.

Norton SJ, Gorga MP, Widen JE, Folsom RC, Sininger Y, Cone-Wesson B, Vohr BR, Fletcher KA. Identification of neonatal hearing impairment: summary and recommendations. *Ear Hear* 2000;21:529–35.

Oyler RF, Lauter JL, Matkin ND. Intrasubject variability in the absolute latency of the auditory brainstem response. *JAAA* 1991;2:206–13.

Parbery-Clark A, Skoe E, Kraus N. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci* 2009;29:14100–7.

Russo N, Nicol T, Musacchia G, Kraus N. Brainstem responses to speech syllables. *Clin Neurophysiol* 2004;115:2021–30.

Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. *Behav Brain Res* 2005;156:95–103.

- Sininger YS. Auditory brain-stem response for objective measures of hearing. *Ear Hear* 1993;14:23–30.
- Sininger YS, Abdala C, Cone-Wesson B. Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. *Hear Res* 1997;104:27–38.
- Skoe E, Kraus N. Auditory brainstem response to complex sounds: a tutorial. *Ear Hear* 2010;31:302–24.
- Smiljanic R, Bradlow AR. Production and perception of clear speech in Croatian and English. *JASA* 2005;118:1677–88.
- Song JH, Banai K, Russo NM, Kraus N. On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurotol* 2006;11:233–41.
- Song JH, Skoe E, Banai K, Kraus N. Perception of speech in noise: neural correlates. *J Cogn Neurosci*, in press.
- Song JH, Skoe E, Wong PCM, Kraus N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 2008;20:1892–902.
- Starr A, Don M. Brain potentials evoked by acoustic stimuli. In: Picton T, editor. *Human event-related potentials. Handbook of electroencephalography and clinical neurophysiology*. New York: Elsevier; 1988. p. 97–158.
- Strait DL, Kraus N, Skoe E, Ashley R. Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci* 2009;29:661–8.
- Swaminathan J, Krishnan A, Gandour JT. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport* 2008;19:1163–7.
- Tice R, Carrell T. *Level 16 (ver. 2.0.3) [computer software]*. Lincoln: University of Nebraska; 1998.
- Tzounopoulos T, Kraus N. Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 2009;62:463–9.
- Van Engen KJ, Bradlow AR. Sentence recognition in native- and foreign-language multi-talker background noise. *JASA* 2007;121:519–26.
- Wechsler D. *Wechsler abbreviated scale of intelligence*. San Antonio, TX: The Psychological Corporation; 1999.