Speech Auditory Brainstem Responses: Effects of Background, Stimulus Duration, Consonant-Vowel, and Number of Epochs

Ghada BinKhamis[1,2], Agnès Léger[1], Steven L. Bell[3], Garreth Prendergast[1], Martin O'Driscoll[1,4], and Karolina Kluk[1]

[1] *Manchester Centre for Audiology and Deafness, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, United Kingdom.*

[2] *King Fahad Medical City, Riyadh, Saudi Arabia*

[3] *Signal Processing and Control Group, Institute of Sound and Vibration Research, University of Southampton, Southampton, UK*

[4] *Manchester Auditory Implant Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom*

Keywords

Speech-ABR, Quiet, Noise, Stimulus Duration, Consonant-vowel

Financial Disclosures/Conflicts of Interest:

This research was funded by the Saudi Arabian Ministry of Education and King Fahad Medical City (to G.B), and by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/M026728/1 (to K.K and S.B).

The authors declare no conflicts of interest.

Portions of this article were previously presented at the XXV International Evoked Response Audiometry Study Group Biennial Symposium, Warsaw, Poland, May 22, 2017, at the 40th MidWinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, USA, February 12, 2017, and at the Basic Auditory Science Meeting, Cambridge, United Kingdom, September 5, 2016.

Address for correspondence:

Ghada BinKhamis, Manchester Centre for Audiology and Deafness (ManCAD), School of Health Sciences, Faculty of Biology Medicine and Health, Room A3.08, Ellen Wilkinson Building, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom. E-mail: ghada.binkhamis@manchester.ac.uk

Karolina Kluk, Manchester Centre for Audiology and Deafness (ManCAD), School of Health Sciences, Faculty of Biology Medicine and Health, Room B2.15, Ellen Wilkinson Building, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom. E-mail: Karolina.Kluk@manchester.ac.uk

Speech Auditory Brainstem Responses: Effects of Background, Stimulus Duration, Consonant-Vowel, and Number of Epochs

ABSTRACT

Objectives: The aims of this study were to systematically explore the effects of stimulus duration, background (quiet versus noise), and 3 consonant-vowels on speech-ABRs. Additionally, the minimum number of epochs required to record speech-ABRs with clearly identifiable waveform components was assessed. The purpose was to evaluate whether shorter duration stimuli could be reliably used to record speech-ABRs both in quiet and in background noise to the 3 consonant-vowels, as opposed to longer duration stimuli that are commonly used in the literature. Shorter duration stimuli and a smaller number of epochs would require shorter test sessions and thus encourage the transition of the speech-ABR from research to clinical practice.

Design: Speech-ABRs in response to 40ms [da], 50ms [ba] [da] [ga], and 170ms [ba] [da] [ga] stimuli were collected from 12 normal-hearing adults with confirmed normal click-ABRs. Monaural (right-ear) speech-ABRs were recorded to all stimuli in quiet, and to 40ms [da], 50ms [ba] [da] [ga], and 170ms [da] in a background of 2-talker babble at +10 dB SNR using a 2-channel electrode montage (Cz-Active, A1 and A2-reference, Fz-ground). Twelve thousand epochs (6000 per polarity) were collected for each stimulus and background from all participants. Latencies and amplitudes of speech-ABR peaks (V, A, D, E, F, O) were compared across backgrounds (quiet and noise) for all stimulus durations, across stimulus durations (50ms and 170ms), and across consonant-vowels ([ba], [da], and [ga]). Additionally, degree of phase locking to the stimulus fundamental frequency (in quiet versus noise) was evaluated for the frequency following response in speech-ABRs to the 170ms [da]. Finally, the number of epochs required for a robust response was evaluated using F_{sp} statistic and bootstrap analysis at different epoch iterations.

Results: *Background effect:* the addition of background noise resulted in speech-ABRs with longer peak latencies and smaller peak amplitudes compared to speech-ABRs in quiet, irrespective of stimulus duration. However, there was no effect of background noise on the degree of phase locking of the frequency following response to the stimulus fundamental frequency in speech-ABRs to the 170ms [da]. *Duration effect:* speech-ABR peak latencies and amplitudes did not differ in response to the 50ms and 170ms stimuli. *Consonant-vowel effect:* different consonant-vowels did not have an effect on speech-ABR peak latencies regardless of stimulus duration. *Number of epochs:* a larger number of epochs was required to record speech-ABRs in noise compared to in quiet, and a smaller number of epochs was required to record speech-ABRs to the 40ms [da] compared to the 170ms [da].

Conclusions: This is the first study that systematically investigated the clinical feasibility of speech-ABRs in terms of stimulus duration, background noise, and number of epochs. Speech-ABRs can be reliably recorded to the 40ms [da] without compromising response quality even when presented in background noise. Since fewer epochs were needed for the 40ms [da], this would be the optimal stimulus for clinical use. Finally, given that there was no effect of consonant-vowel on speech-ABR peak latencies, there is no evidence that speech-ABRs are suitable for assessing auditory discrimination of the stimuli used.

INTRODUCTION

The Auditory Brainstem Response (ABR) is an auditory evoked potential that is recorded from the scalp in response to multiple short auditory stimuli such as clicks, tone bursts, or chirps (Hall 2015). The ABR to clicks and tone-bursts is a well-established clinical measure that is widely used to evaluate hearing in patients that are unable to perform standard behavioral hearing threshold measures. The ABR has advantages over other auditory evoked potentials in that it is not influenced by attention or state of arousal, and that the response can be reliably recorded from infants and young children (Hall 2015; Hood 2015). The ABR could also be measured in response to short consonant vowel (CV) stimuli (e.g., [ba] [da] [ga]) (Skoe & Kraus 2010). This type of ABR will be referred to as the speech-ABR. It has been shown that the speech-ABR waveform follows the temporal and spectral features of the CV stimulus, these features play an important role in speech understanding in that: (i) onset of sound facilitates phoneme identification; (ii) frequency transitions allow consonant identification; (iii) formant structure facilitates vowel identification; and (iv) the fundamental frequency (F_0) portrays non-linguistic information such as gender and emotion (Kraus & Nicol 2005; Abrams & Kraus 2015). These temporal and spectral features of speech cannot be measured through current clinical ABRs to click and tone-burst stimuli. It has therefore been proposed that the speech-ABR may be used as a measure of: (i) brainstem speech encoding (e.g., Kraus & Nicol 2005; Johnson et al. 2005; Chandrasekaran & Kraus 2010); (ii) speech-in-noise performance, where responses in noise are more degraded with longer peak latencies and smaller peak amplitudes than responses in quiet, and are more degraded in individuals who perform worse on behavioral speech-in-noise measures compared to those who perform better (e.g., Anderson et al. 2011; Parbery-Clark et al. 2011; Song et al. 2011b); and (iii) auditory discrimination of different CVs, where CVs with a higher second formant (F_2) frequency have shorter peak latencies than CVs with a lower F_2 frequency (e.g., Johnson et al. 2008; Hornickel et al. 2009b). Therefore, the speech-ABR may have potential for clinical application in audiology as an objective measure of detection of speech sounds, speech-in-noise performance, and discrimination of different speech sounds. The speech-ABR could compliment currently available clinical ABRs that were introduced into audiology clinical practice in the 1980s (Galambos & Despland 1980) following a period of lab-based investigations since the discovery of ABRs in 1970 (Jewett et al. 1970). The reader is referred to Hall 2015 (chapters 4 and 5) for a review of the transition of current clinical ABRs from research to clinical practice.

The length of CV stimuli used in the literature ranges from short (no sustained vowel period), to long, e.g. 40ms (e.g., Hornickel et al. 2009a; Krizman et al. 2010), 60ms (e.g., Akhoun et al. 2008), 170ms (e.g., Johnson et al. 2008; Song et al. 2011b), and 180ms (e.g., Bellier et al. 2015). The shorter CV (40ms) contains an onset burst and a formant transition period without a sustained vowel period. Subsequently, speech-ABRs to the 40ms [da] contain onset peaks (V and A), transition peaks (D, E, and F), and offset peak (O) (e.g., Hornickel et al. 2009a). The longer CVs (170 and 180ms) contain an onset burst, a formants transition period, and a sustained vowel period. Subsequently, speech-ABRs to longer CVs contain onset and transition peaks and an additional frequency following response (FFR) (e.g., Johnson et al. 2008; Bellier et al. 2015). Researchers who used the speech-ABR to assess speech-in-noise performance mainly used the 170ms [da] (e.g., Anderson et al. 2011; Parbery-Clark et al. 2011; Song et al. 2011a,b; Hornickel et al. 2012), while the 40ms [da] was used only by a few (e.g., Russo et al. 2004; Anderson et al. 2013a). Additionally, 170ms [ba] [da] [ga] were researched in the context of evaluating discrimination between CVs via the speech-ABR (e.g., Johnson et al. 2008; Hornickel et al.

2009b). The rationale behind selecting longer stimuli over shorter stimuli for speech-ABRs in noise and for speech-ABRs to different CVs has not been discussed in the literature. While the use of longer stimuli that contain a sustained vowel period or a vowel with changing pitch trajectories would be necessary to assess certain populations such as native speakers of tonal languages (e.g., Krishnan et al. 2005; Swaminathan et al. 2008) or individuals with autism spectrum disorder (e.g., Russo et al. 2008), shorter stimuli may be appropriate to elicit speech-ABRs in noise and speech-ABRs to different CVs. We postulate that longer stimuli are commonly used because they have a closer resemblance to natural speech and their speech-ABRs contain a sustained period (FFR) that would result in responses that contain more components than responses to shorter stimuli. Although longer stimuli would require longer recording sessions, which may hinder the speech-ABRs' clinical applicability. Nonetheless, the effect of stimulus duration on the speech-ABR in noise and the speech-ABR to different CVs has not yet been assessed.

The speech-ABR has the potential to become a clinical audiological measure. However, stimulus duration would influence the implementation of the speech-ABR in the clinical setting. Specifically, shorter stimuli would be more clinically feasible as they would require shorter recording sessions. Shorter stimuli have been used to record speech-ABRs in noise and thus may have potential use in assessing speech-in-noise performance with the speech-ABR (e.g., Russo et al. 2004; Anderson et al. 2013a). With regards to the use of speech-ABRs to assess discrimination between CVs, shorter stimuli may be sufficient to record speech-ABRs if the difference in F_2 frequency between CVs is reflected in the vowel formant transition period for each CV. Additionally, the method used to analyze discrimination between CVs should not require the sustained period as a control condition as is required in cross-phasogram analysis (e.g., Skoe et al. 2011). Another factor that may influence the clinical implementation of the speech-ABR is the minimum number of epochs (number of repetitions) required to obtain a response with clearly identifiable waveform components (peaks). A larger number of epochs requires longer recording sessions. Number of epochs used in speech-ABR literature ranges from 4000 to 6000 (e.g., Johnson et al. 2008; Hornickel et al. 2009a; Skoe & Kraus 2010; Skoe et al. 2015). However, the minimum number of epochs required to obtain speech-ABRs with clearly identifiable peaks has not yet been addressed.

The aim of this study was to assess the effect of background (quiet versus noise) and stimulus duration on speech-ABRs. Speech-ABR time domain waveforms evoked by 3 CVs ([ba] [da] [ga]) of short duration (40ms and 50ms) and long duration (170ms) in 2 backgrounds (quiet and noise) were evaluated in order to: (i) assess if short CVs can be reliably used to measure speech-ABRs in quiet and in noise; (ii) evaluate the differences in responses to short versus long CVs; and (iii) determine if auditory discrimination between CVs ([ba], [da], [ga]) can be assessed with short CVs. The issue of the minimum number of epochs required to obtain a speech-ABR with clearly identifiable peaks was also addressed.

MATERIALS AND METHODS

Participants

Twelve adults (age 22 – 49 years, mean = 31.42, SD = 7.88, 7 females) with normal hearing ($\leq$25 dBHL at 250 – 8000Hz), normal click-ABRs at 100 dB peak equivalent SPL (peak latencies (ms); I: mean = 1.86, SD: 0.18, III: mean = 4.00, SD = 0.19, V: mean = 5.89, SD = 0.21), and no history of neurological disorders or learning difficulties were tested. Participants were recruited from the University of Manchester and were compensated for their time. All participants provided written informed consent before enrolment in this study.

This study was approved by the University of Manchester research ethics committee (Ref: UREC 15487).

Speech-ABR recording

- Equipment

Raw EEG responses were collected with Cambridge Electronic Design (CED, Cambridge, UK) 'Signal' software (Version 5.11) using a CED power 1401 mkII data acquisition interface (CED Limited) and a Digitimer 360 isolated 8-channel patient amplifier (Digitimer Limited, Hertfordshire, UK). Speech-ABR stimuli were presented from the CED 'Signal' software through the CED power 1401 mkII and routed through a Tucker-Davis Technologies (TDT, Alachua, FL, USA) PA5 Programmable attenuator and a TDT HB7 Headphone Driver to E.A.RTONE 3A insert earphones (E.A.R Auditory Systems, Aearo Company, Indianapolis, IN, USA). Background noise was presented from Audacity (version 1.2.6) via an E-MU 0202 sound card (Creative Technology Limited, UK) and routed through the TDT HB7 Headphone Driver to the E.A.RTONE 3A insert earphones; splitters were used in order for the stimuli and noise to be presented through the same insert earphone. Stimuli (CVs and background noise) were calibrated in dB A using a Brüel and Kjær type 2250 (Brüel and Kjær, Nærum, Denmark) sound level meter.

- Stimuli

Three stimulus durations were used: (i) 5-formant synthesized 40ms [da] (described in Banai et al. 2009); (ii) 6-formant synthesized 50ms [ba] [da] and [ga]; and (iii) 6-formant synthesized 170ms [ba] [da] and [ga] (described in Hornickel et al. 2009b). The 40ms [da] and the 170ms

CVs ([ba] [da] [ga]) are identical to those used in the literature; however, the 50ms CVs ([ba] [da] [ga]) are not, but they are identical to the first 50ms of the 170ms CVs. The 170ms CVs differed in the frequency of F_2 during the formant transition period with F_0 and other formant frequencies equal across the 3 CVs. The 50ms CVs were created by clipping the 170ms CVs at the end of the formant transition period (50ms) using hamming windowing in MATLAB (Version R2015a, MathWorks). The first 40ms of each CV was kept unaltered and > 90% reduction in amplitude was applied over the last 10ms. The resulting 50ms [ba] [da] and [ga] contained the onset burst and transition period of the original 170ms CVs without the sustained period. The 40ms [da] stimulus differed from the 50ms and 170ms CVs in that it contained a longer onset burst and only 5 formants as opposed to the 6 formants in the other CVs (see document, Supplemental Digital Content 1, Section 1: Characteristics of CV Stimuli). Polarity of all CVs was reversed using Adobe Audition CC (2015.1 Release, build 8.1.0.162) in order to evoke speech-ABRs using 2 opposite stimulus polarities as recommended by Skoe and Kraus (2010).

Speech-ABRs in noise were measured using a 2-talker-babble masker (used by Song et al. 2011a,b). Two-talker babble was selected over speech spectrum noise as being more representative of real life situations, and to ensure that the ABR in noise fell between celling (response in quiet) and floor (EEG noise floor). Since 2-talker babble contains deep modulations, it degrades the speech-ABR less than the 6-talker babble as shown by Song et al. (2011b). Speech-ABRs in 2-talker babble have been previously described in response to the 170ms [da] (e.g Song et al. 2011b); however, to our knowledge, this is the first study to describe speech-ABRs to the 40ms [da] in 2-talker babble.

- Recording Parameters

CED ‘Signal’ software sampling configuration was set to gap-free sweep mode, sample rate of 20000 Hz, pulses with a resolution of 0.01ms as the output type, and outputs were set at absolute levels and absolute times. Online artifact rejection was set to reject epochs that included any activity above 20µV. Stimulus presentation rates were stimulus specific and were set based on the stimulus duration plus an inter-stimulus interval sufficient to record the response and the baseline (Skoe & Kraus 2010). Since recording time would influence the clinical applicability of the speech-ABR, presentation rates were therefore set to reduce recording time to the shortest possible per stimulus (See Table 1 for additional parameters). Two channel vertical electrode montage recording with Cz active, earlobe reference (A1 and A2), and high forehead ground (Fz) was conducted, electrode sites were based on the international 10-20 EEG system.

Procedure

- Participant Preparation

Skin at Cz, earlobes (A1 and A2), and high forehead (Fz) was prepared using Nuprep Skin Prep Gel. Ag/AgCI 10mm disposable disc electrodes were placed on prepared sites with Ten20 Conductive EEG paste and secured with tape at A1, A2, and Fz.

- Recording Environment

Participants were seated and reclined in a comfortable recliner in a double-wall soundproof booth, and instructed to remain relaxed with their eyes closed in order to reduce myogenic artifacts and eye blinks. Insert earphone was placed in the right ear with the appropriate sized E.A.RLINK foam ear-tip while the left ear remained free. Right ear recording was selected due to the reported right ear advantage for speech-ABR (Hornickel et al. 2009a).

- Recording Sessions

Speech-ABRs in quiet were collected in response to the 40ms [da], 50ms [ba] [da] [ga], and 170ms [ba] [da] [ga]. Speech-ABRs in 2-talker babble at +10 dB signal to noise ratio (SNR) (70 dB A noise and 80 dB A speech) were collected in response to the 40ms [da], 50ms [ba] [da] [ga], and only the 170ms [da]. SNR of +10 dB was set based on speech-ABR literature. Background babble was paused after each block and restarted at the next block to ensure random sections of the babble started with each block. Recordings were completed over 4 to 5 sessions (2 to 3 hours each) across 4 to 5 weeks. Order of the 2 backgrounds (quiet and noise) and order of CVs and durations were randomized using a Latin square. A total of 12000 artifact free epochs were collected per stimulus, 2 blocks of 3000 epochs were collected for each stimulus polarity resulting in a total of 6000 epochs per polarity. Electrode impedances were below 3 kΩ and impedances between electrodes were balanced and below 1 kΩ. Recording times were documented from the start of the first block until the end of the fourth block per stimulus and background (quiet and noise); including rejected epochs and repeated blocks due to increased EEG artifact. Recording times for speech-ABRs to the 40ms [da] were slightly shorter than to the 50ms CVs. Speech-ABRs to the 170ms CVs took longest (see document, Supplemental Digital Content 1, Section 2: Recording Time Per Stimulus).

Analyses

- Processing ABRs

Raw EEG data were processed and analyzed in MATLAB R2015a (MathWorks). The ipsilateral channel (channel 2) was processed for each response. The 2 blocks of each polarity were averaged separately then low-pass filtered at 2000Hz as reported in the speech-ABR literature (e.g., Russo et al. 2004; Banai et al. 2009; Anderson et al. 2013b), using the *eegfilt* function of the EEGLAB toolbox (Delorme & Makeig 2004). Filtered responses for each polarity were then averaged together for a final averaged alternating polarity response. Alternating polarity was used in order to reduce stimulus artifact and cochlear microphonics (Skoe & Kraus 2010). Final responses were then baseline corrected via de-meaning and the first 70ms were plotted in the time domain to assess peak latencies and peak amplitudes. Time domain analyses were preferred in order to maintain clinical applicability. Although other analyses techniques are emerging and clinical practice may change in the future, to date clinical audiologists analyze click and tone burst ABRs in the time domain. Final high-pass filter setting (70 Hz) used for the [ba] [da] [ga] CVs in this study was different than the setting (300 Hz) used by Johnson et al. (2008) and Hornickel et al. (2009b). Johnson et al. and Hornickel et al. reported initially high-pass filtering at 70 Hz, then applying an additional high-pass filter of 300 Hz to emphasize the differences in peak latencies between [ba] [da] and [ga]. However, speech-ABRs recorded for this study were obliterated when high-pass filter was set to 300 Hz, therefore speech-ABR major and minor peaks identified by Johnson et al. and Hornickel et al. could not be identified in this study (see document, Supplemental Digital Content 1, Section 3: Filtering Speech-ABRs to Emphasize Peak Latency Differences Between [ba], [da], and [ga], Section 4: Why Speech-ABRs Contained No Spectral Peaks Above 300 Hz). Thus all results presented for the [ba] [da] [ga] CVs below were high-pass filtered at 70 Hz.

- Peak latency and Amplitude Measurements

To account for the length of the tube of the E.A.RTONE 3A insert earphones, the value of 0.8ms was subtracted from each peak latency value (Van Campen et al. 1992). Positive peak V and negative peaks A, D, E, F, and O that have been reported in the 40ms [da] speech-ABR literature (e.g., Skoe & Kraus 2010; Skoe et al. 2015) were visually identified based on published peak latency normative data (Skoe et al. 2015) and their latencies were measured for the 40ms [da] speech-ABRs. For the 50ms and 170ms CVs, peaks that corresponded to the 40ms [da] peaks in terms of peak latency and order of occurrence in the response were visually identified and their latencies were measured. In order to remain consistent, the same peak nomenclature was used for responses to the 50ms and 170ms CVs. Thus, peak O in response to the 40ms and 50ms CVs is an offset peak, but it is an early FFR peak in response to the 170ms CVs. Peak (V) to trough (A) amplitudes were measured. For negative peaks D, E, F, and O, the positive peak preceding each peak was used for peak to trough amplitude measurements.

- Verifying speech-ABR quality and identified peaks

Two methods were used to assess quality of responses and ensure 95% confidence that visually identified peaks were above the EEG noise floor. First, the F_{SP} statistic was applied with a criterion of $F_{SP} \geq 3.1$ (as described by Don et al. 1984; Elberling & Don 1984). F_{SP} is a measure of the variance in the response over the variance in the background EEG noise, measured by comparing the EEG data within a time region where the response is expected to occur (variance in the response) to the variance of the EEG data at a single time point (variance in the EEG background noise) across averaged epochs (Don et al. 1984; Elberling & Don 1984). Elberling and Don (1984) reported that an F_{SP} of 3.1 equated to 99% confidence that their click-ABRs were present and above the EEG noise floor, and this was measured based on what they termed as a "worst case" (i.e., participants with the highest variance in their background EEG noise). The criterion of $F_{SP} \geq 3.1$ set for this study was informed by the work by Don et al. (1984) and Elberling and Don (1984) on click-ABRs as there is no literature on F_{SP} and speech-ABRs. This was applied with the knowledge that there may be individual variability between participants depending on their background EEG noise, differences in filter settings used in this study compared to those used by Don et al. and Elberling and Don, and differences in stimuli (CVs versus clicks). F_{SP} analyses time windows were: 5 – 60ms for responses to 40ms [da], 8 – 70ms for responses to both the 50ms and 170ms stimuli. The position of F_{SP} single point was set in the middle of each time window specified above. Speech-ABRs in quiet were considered present if $F_{SP} \geq 3.1$. F_{SP} was measured for speech-ABRs in noise; however, since F_{SP} literature only reported results from testing in quiet and there has not been criterion reported for testing in noise, the criterion of 3.1 was not applied to speech-ABRs in noise. Additionally, speech-ABRs in noise have been shown to have lower SNRs compared to speech-ABRs in quiet (Song et al. 2011a; Hornickel et al. 2012); therefore, it is likely that F_{SP} values will also be lower. F_{SP} was measured to no sound recordings and F_{SP} values were <1.5 (mean = 0.95, SD = 0.25) for all participants, and F_{SP} values of speech-ABRs in noise that did not reach 3.1 were all >1.7 (mean = 2.67, SD = 0.45). Therefore, speech-ABRs in noise were considered present when the F_{SP} at 12000 epochs was above the participants' "no sound" F_{SP}. Second, the bootstrap method (Efron 1979a,b; Efron 1981); a method that estimates confidence intervals; was applied (as described by Lv et al. 2007). The bootstrap method does not rely on the variability between participants and can estimate the significance of F_{SP} values for each individual recording. Bootstrap was used to confirm that visually identified peaks were with 95% confidence above the noise floor (Fig. 1), any visually identified peaks that fell outside the 95% confidence lines were considered absent. Both F_{SP} and bootstrap were applied to the 12000-epochs of speech-ABRs evoked by all stimuli.

- Determining Number of Epochs Required for a Robust Response

F_{SP} and bootstrap were used to evaluate the number of epochs required to record speech-ABRs with clearly identifiable peaks in response to the 40ms and 170ms [da] in quiet and in noise. Both methods were applied to the averaged alternating polarity speech-ABRs at 15 iterations starting at 800 epochs and increasing by 800 up to 12000 epochs. The first criterion was the minimum number of epochs required to reach an $F_{SP} \geq 3.1$. Once this value was reached, the number of epochs (at or above the number required for $F_{SP} \geq 3.1$.) required for all speech-ABR peaks that were detected at 12000 epochs to be detected with 95% confidence via bootstrap were evaluated for each participant.

- Degree of FFR Phase Locking

To assess the effect of background noise on the FFR, inter-trial phase clustering (degree of phase locking) to F_0 of the stimulus was implemented on the FFR period (70 – 190ms) of the raw EEG responses to the 170ms [da] in quiet and in noise using the method recommended by Cohen (2014). Inter trial phase clustering is the length of the average vector measured by extracting the phase angle for a specific frequency (F_0 in this study) at each time point from each epoch, and calculating the average vector length from the distribution of phase angles in a polar plane, resulting in a value between 0 and 1. Values closer to 1 indicate similar phase angles and thus a higher degree of phase locking, and values closer to 0 indicate minimal degree of phase locking at a particular time point (Cohen 2014). Phase locking analyses focused on F_0 as it was the most robust component present in speech-ABRs of all participants.

- Statistical Analyses
    - Effect of Background

The effect of background (quiet versus noise) on peak latencies and peak amplitudes of speech-ABR peaks (V, A, D, E, F, O) was evaluated through fitting linear mixed models (LMM) in **R** (R Core Team 2016) using *lemer* of the *lme4* package (Bates et al. 2015) and *lmerTest* (Kuznetsova et al. 2016). LMMs allow for unbalanced designs and account for missing data points (e.g., missing peaks in some participants). Two LMMs were fit to the data: (1) latency model was set up with 'background' (quiet and noise), 'duration' (40ms, 50ms, and 170ms), 'peak' (V, A, D, E, F, O), and interaction between ‘duration’ and ‘peak’ as fixed effects and ‘participants’ as random effects, (2) amplitude model was set up with ‘background’, ‘duration’, ‘peak’, interaction between ‘background’ and ‘peak’, interaction between ‘background’ and ‘duration’, and interaction between ‘duration’ and ‘peak’ as fixed effects and ‘participants’ as random effects. LMMs were built by conducting a likelihood ratio test to compare an LMM with a fixed effect to a LMM without the fixed effect as described by Winter (2013). Fixed effects that had a significant effect on the LMM ($p < 0.05$) plus LMMs that resulted in a better fit to the data in terms of lower Akaike’s information criterion (AIC) were finally selected. More complex LMMs with random intercepts were attempted; however, these models did not converge. The LMM without ‘CV’ ([ba], [da], [ga]) as a fixed effect was a better fit to the data; therefore, ‘CV’ was dropped as a fixed effect from both latency and amplitude models.

Next, the effect of background on the FFR period of the speech-ABR to the 170ms [da] was evaluated by conducting a 2-tailed paired sample *t* test using **R** on the Fisher-Z transformed maximum degree of phase locking to the fundamental frequency (F_0) in quiet versus in noise.

- Effect of Stimulus Duration:

The effect of stimulus duration on peak latencies and peak amplitudes of speech-ABR peaks (V, A, D, E, F, O) was evaluated via conducting 2 LMMs that were the best fit to the data: (1) latency model was set up with ‘background’ (quiet and noise), ‘duration’ (50ms, and 170ms), ‘peak’ (V, A, D, E, F, O), and interaction between ‘duration’ and ‘peak’ as fixed effects and ‘participants’ as random effects, (2) amplitude model was set up with ‘background’, ‘duration’, ‘peak’, and interaction between ‘duration’ and ‘peak’ as fixed effects and ‘participants’ as random effects. The duration comparison was restricted to the 50ms and 170ms CVs and the 40ms [da] was excluded due to the spectral differences in the stimulus that may influence results.

- Effect of CV:

In order to evaluate the effect of CV on peak latencies, a simpler LMM latency model was built using only speech-ABRs in quiet to 50ms and 170ms [ba] [da] [ga], with 'CV' and interaction between 'peak' and 'CV' as fixed effects and 'participants' as random effects.

All post hoc pairwise comparisons were conducted using the *lsmeans* (Lenth 2016) **R** package. Bonferroni correction was applied to all *p* values to correct for multiple comparisons. Criteria for significance was considered $p < 0.01$.

RESULTS

Detected Peaks

Most peaks were detected with 95% confidence via bootstrap in speech-ABRs of all participants in quiet, with more peaks missing in speech-ABRs in noise than in quiet (see document, Supplemental Digital Content 2, Section 1: Detection of Speech-ABR Peaks). The most commonly missing peak was V in noise in speech-ABRs to all stimuli excluding the 40ms [da], where F was the peak most commonly missing in speech-ABRs in noise.

Effect of Background

- Peak latencies

Background had a significant effect on speech-ABR peak latencies ($b = 0.91$, $t(796.10) = 9.42$, $p < 0.01$), (Figs. 2 and 3). Peak latencies in noise were longer than peak latencies in quiet for all stimulus durations. Post hoc pairwise comparisons to investigate the effect of 'background' on specific peak latencies revealed that latencies of all peaks were significantly longer ($p < 0.01$) in speech-ABRs in noise compared to in quiet regardless of stimulus duration (see document, Supplemental Digital Content 2, Section 2: Speech-ABR Mean (SD) Peak Latencies and

Amplitudes, Section 3: Effects of Background on Speech-ABRs – Post Hoc Pairwise Comparison Results).

- Peak amplitudes

Peak amplitudes for speech-ABRs in noise were significantly smaller than peak amplitudes in quiet (b = -0.12, $t(687.00) = -6.24$, $p < 0.01$) (Figs. 2 and 3). There was a significant interaction between 'background' and 'peak' ($\chi^2(1) = 30.09$, $p < 0.01$) as revealed by the likelihood ratio test. Post hoc pairwise comparisons to investigate the effect of 'background' on specific peak amplitudes revealed that all speech-ABR peaks had significantly smaller amplitudes ($p < 0.01$) in noise compared to in quiet regardless of stimulus duration, excluding peak O that had a similar amplitude in quiet and in noise (see document, Supplemental Digital Content 2, Section 2: Speech-ABR Mean (SD) Peak Latencies and Amplitudes, Section 3: Effects of Background on Speech-ABRs – Post Hoc Pairwise Comparison Results).

- Degree of phase locking

Greater FFR degree of phase locking to F_0 was found in speech-ABRs in noise relative to in quiet (Fig. 4), though this difference was not significant ($t(21.97) = -0.29$, $p = 0.78$).

Effect of Stimulus Duration

- Peak latencies

Stimulus duration (50ms versus 170ms) did not have a significant effect on speech-ABR peak latencies (b = 0.74, $t(667) = -2.815$, $p = 0.09$) (Fig. 5).

- Peak amplitudes

Peak amplitudes for speech-ABRs to the 50ms CVs were significantly smaller than to 170ms CVs (b = -0.07, $t(578) = -3.83$, $p < 0.01$) (Fig. 5). There was a significant interaction between 'duration' and 'peak' ($\chi^2(1) = 18.46$, $p < 0.01$) as revealed by the likelihood ratio test. Post hoc pairwise comparisons to investigate the effect of 'duration' and 'the interaction between 'duration' and 'peak' on specific peak amplitudes revealed that only peak D amplitude was significantly smaller ($p < 0.01$) in speech-ABRs to the 50ms CVs compared to the 170ms CVs both in quiet and in background noise.

Effect of CV

'CV' had no effect on peak latencies ([da]: $b = -0.04$, $t(396) = -0.10$, $p = 0.92$, [ga]: $b = -0.01$, $t(396) = -0.10$, $p = 0.99$)(Fig. 5); however, there was a significant interaction between 'peak' and 'CV' ($\chi^2(1) = 2201.90$, $p < 0.01$) as revealed by the likelihood ratio test. Post hoc pairwise comparison to investigate this interaction revealed no significant effect of 'CV' on peak latencies when comparison was on the same peak and a different CV (e.g., peak D and CV [ba] versus peak D and CV [ga]). Some authors (e.g., Skoe et al. 2011) have suggested using a 'cross-phaseogram' approach to explore how the phase of components in speech-ABRs to different CVs may vary. This approach uses the cross-power spectral density between the responses to 2 CVs to calculate phase differences between the responses over time and frequency. Use of this approach for analyses of speech-ABRs from this study was not appropriate due to the following: (i) phase measurements are very sensitive to background noise and this generally increases when responses are combined; (ii) the analyses will include frequencies that are not harmonics of the fundamental frequency in the response and hence phase would be calculated at frequencies where no response would be expected, which introduces difficulty in interpretation; and (iii) the robustness and efficacy of the cross-phaseogram has not yet been well tested.

Number of Epochs

The numbers of epochs required to reach $F_{SP} \geq 3.1$ varied among participants, which may reflect variations in the background EEG noise characteristics between participants. In general, speech-

ABRs in quiet required a smaller number of epochs to reach $F_{SP} \geq 3.1$ than speech-ABRs in noise to both 40ms and 170ms [da]. In 2 participants, speech-ABRs in noise to the 170ms [da] did not reach $F_{SP} \geq 3.1$ ($F_{SP} = 2.96$, $F_{SP} = 2.95$) at 12000 epochs; however, their speech-ABR peaks were detected with 95% confidence via bootstrap. Although criterion of $F_{SP} \geq 3.1$ indicates that response is present, it does not imply that all peaks can be detected, as some participants required more epochs for all peaks to be detected with 95% confidence via bootstrap than to reach $F_{SP} \geq 3.1$. Specifically, in speech-ABRs to the 40ms [da], 4 participants required 800 more epochs in order for all peaks to be detected in their speech-ABRs in quiet, and 5 participants required a larger number of epochs (1 required 800, 2 required 1600, and 2 required 4000 more epochs) for all peaks to be detected in their speech-ABRs in noise. In speech-ABRs to the 170ms [da], 7 participants required larger number of epochs (1 required 1600, 2 required 2400, 2 required 3200, 1 required 4000, and 1 required 4800 more epochs) for all peaks to be detected in their speech-ABRs in quiet, and 5 participants required larger number of epochs (2 required 800, 2 required 4000, 1 required 2400, and 2 required 4800 more epochs) for all peaks to be detected in their speech-ABRs in noise (see document, Supplemental Digital Content 2, Section 4: Bootstrap Results and Examples). Average F_{SP} values where all peaks were detected with 95% confidence via bootstrap for speech-ABRs in quiet were 4.17 (SD = 0.91, range: 3.16 – 6.17) for the 40ms [da] and 6.94 (SD = 3.65, range: 3.25 – 12.42) for the 170ms [da], and for speech-ABRs in noise were 4.24 (SD = 1.15, range: 3.14 – 6.16) for the 40ms [da] and 4.30 (SD = 1.82, range: 3.21 – 8.86) for the 170ms [da] (see document, Supplemental Digital Content 2, Section 5: Fsp Values and Number of Epochs).

DISCUSSION

The aims of this study were to evaluate the effects of: background, stimulus duration, and CV on speech-ABRs. Hence, the differences in speech-ABRs recorded to 3 CVs of short duration (40 and 50ms) and long duration (170ms) presented in 2 backgrounds (quiet and noise) were assessed. This was done in order to establish if shorter CVs, that would be more clinically applicable due to shorter test-time: i) can be reliably used for speech-ABRs in noise; ii) evoke robust ABRs comparable to ABRs evoked by long CVs; and iii) can be used to assess discrimination between CVs. A secondary aim of this study was to evaluate the number of epochs required to achieve a speech-ABR with clearly identifiable peaks. It is worth noting that results from this study apply to recording speech-ABRs at 80 dB A, and response quality may be reduced if lower presentation levels are to be used.

Speech-ABR in Background Noise

Speech-ABR peak latencies were longer and amplitudes were smaller in noise than in quiet across the 3 durations and the 3 CVs, excluding amplitude of peak O that was not affected by background noise. Additionally, there were more speech-ABR peaks missing in noise than in quiet. These results are in general agreement with published results on speech-ABRs in noise for the 40ms and 170ms [da] (Russo et al. 2004; Parbery-Clark et al. 2011; Song et al. 2011a). Results are also in agreement with published results on click-ABRs in noise that found a delay in click-ABR peak V (analogous to speech-ABR peak V) latency when background noise was added (e.g., Burkard & Sims 2002; Mehraei et al. 2016). However, Parbery-Clark et al. (2011) reported that only onset peaks had reduced amplitudes in noise compared to in quiet, with longer latencies of both onset and transition peaks in noise, and Song et al. (2011a) reported that only onset peaks V and A had delayed latencies with no difference in latencies of transition peaks between quiet and noise. Parbery-Clark et al. recorded speech-ABRs binaurally to the 170ms [da], binaural presentation is known to result in more robust responses (Skoe & Kraus 2010), which may explain the lack of change in amplitudes in transition peaks found by Parbery-Clark et al.. While there were no notable methodological differences between this study and Song et al.. The reasons behind our longer peak latencies and smaller peak amplitudes in noise compared to in quiet are unclear. Burkard and Sims attributed click-ABR peak V latency delay to neural desynchronization. Mehraei et al. also stipulated that neural desynchronization resulted in delayed click-ABR peak V latency, more specifically that low spontaneous rate auditory nerve fibers that are slower to fire are the main contributors to ABRs in noise, while high spontaneous rate auditory nerve fibers contribute less because they are more affected by background noise. Another reason may be that the addition of background noise may result in a shift in cochlear-place of the response, as it has been shown that speech-ABRs in quiet that originated from a lower-frequency cochlear region had longer peak latencies and smaller peak amplitudes (Nuttall et al. 2015). Furthermore, the lack of difference in peak O amplitudes in noise compared to in quiet may be a result of compensation that occurs in the brainstem pathway as stipulated by Russo et al. (2004). In terms of the effect of background on the FFRs degree of phase locking to F_0 of the stimulus, we found no significant difference between speech-ABR FFRs in quiet and in noise. This lack of effect of background noise on F_0 is consistent with earlier reports (Li & Jeng 2011; Song et al. 2011b; Smalt et al. 2012). Li and Jeng (2011) also found that F_0 of the FFR did not decrease in amplitude with positive dB SNR levels, it was only affected at 0 dB SNR and negative dB SNR levels. While AlOsman et al. 2016 and Prévost et al. 2013 found an enhancement in FFR F_0 in background noise compared to in quiet. AlOsman et al. stipulated that this enhancement was modulated by top down processing in order to improve speech understanding in background noise, while Prévost et al. attributed this enhancement to the phase locking to the stimulus envelope of auditory nerve fibers that are further away from the characteristic frequency of F_0, in order to compensate for the effect of background noise. Involvement of the auditory cortex in the FFR has been shown by Coffey et al. (2016) in their FFR and Magnetoencephalography (MEG) study where auditory cortical activation at F_0 of the stimulus was found in normal hearing adults. This supports top down modulation of the FFR and may explain the lack of effect of background noise on phase locking to F_0 that was found in this study. However, a significant effect of background on peak latencies and amplitudes occurring in the first 60–70ms of the speech-ABR was found. Physiological reasons behind these effects remain unclear as physiological mechanisms related to speech perception in noise within the peripheral auditory system and the brainstem are still not fully resolved in the literature. Further investigation of these physiological mechanisms is needed. Nonetheless, the effect of background noise on speech-ABR peak latencies and amplitudes was similar across the 3 CV durations in this study, and the FFR period (70 – 190ms) of the speech-ABR to the longer duration stimulus was not affected by background noise at +10 dB SNR. These results suggest that peaks occurring in the first 60–70ms's of speech-ABRs to all stimulus durations are equally influenced by noise with the FFR period to the longer stimulus durations not being affected by noise. The FFR period would likely require higher background noise levels in order to be affected, which would require higher presentation levels that may be uncomfortably loud to some individuals as was revealed during the pilot for this study.

Speech-ABRs and Stimulus Duration

Speech-ABR peak latencies and peak amplitudes were similar across the 50ms and 170ms CVs. Although faster presentation rates have been reported to delay onset peak latencies (Krizman et al. 2010), this was not the case in this study. Peak latencies of speech-ABRs to 170ms CVs (presented at 4.35 stimuli per second) were similar to those in response to the 50ms CVs (presented at 9.1 stimuli per second). These results suggest that stimulus duration does not affect speech-ABR peak latencies or peak amplitudes when shorter and longer versions of the same stimuli are used, and all speech-ABR peaks are identifiable across the 2 durations (50ms, 170ms). Therefore, any stimulus duration may be used to record speech-ABRs, assuming stimulus specific normative data is established.

Speech-ABR and CV Discrimination

Speech-ABR peak latencies to the 3 CVs ([ba] [da] [ga]) were similar across the 3 CVs and 2 durations (50ms and 170ms) in quiet. These results are at odds with results from Johnson et al. (2008) and Hornickel et al. (2009b) who found overall earlier peak latencies for the 170ms [ga] compared to the 170ms [da] and [ba], and overall later peak latencies for the 170ms [ba] compared to the 170ms [da] and [ga]. Speech-ABR high-pass filter cut-off frequency used by Johnson et al. and Hornickel et al. was 300 Hz. High-pass filtering speech-ABRs from this study at such a high frequency resulted in complete loss of the response, thus the major and minor peaks that were identified by Johnson et al. and Hornickel et al. could not be identified in speech-ABRs from this study. The reasons behind differences between speech-ABRs recorded in this study and those recorded by Johnson et al. and Hornickel et al. are unclear. Speech-ABRs from this study contained little to no spectral peaks above 300 Hz, which rendered high-pass filtering at 300 Hz redundant. Additionally, the spectra of speech-ABRs from this study were a very good match to the predicted spectra obtained from analyzing the half-wave rectified acoustic CV stimuli (these same CVs were used in Johnson et al. 2008; Hornickel et al. 2009b) that also contained no clear spectral peaks above 300 Hz. Therefore, it is unclear what is driving the high frequency content in speech-ABRs reported by Johnson et al. and Hornickel et al.. Also, the 3 CVs only differ in the vowel formant frequency of F_2, which is above the reported maximum frequency (approximately 1034 Hz) that the brainstem is able to phase-lock to (Liu et al. 2006). Results from this study suggest that the speech-ABR may not be a useful tool to assess auditory discrimination between these specific CVs that differ in F_2 frequency regardless of CV duration.

Number of Epochs

Number of epochs required for recording speech-ABRs with clearly identifiable peaks varied between participants, they were as low as 1600 in quiet and 2400 in noise and as high as 6400 in quiet and 12000 in noise. Number of epochs required for speech-ABRs in noise was generally larger than in quiet to both the 40ms and 170ms [da]. Speech-ABRs to the 40ms [da] in quiet and in noise of most participants (total = 8) required a smaller number of epochs to reach a combination of $F_{SP} \geq 3.1$ and peaks detected with 95% confidence via bootstrap than to the 170ms [da]. Plus, speech-ABRs in noise to 170ms [da] did not reach $F_{SP} \geq 3.1$ at 12000 epochs in 2 participants. Fewer epochs to achieve speech-ABRs with clearly identifiable peaks in response to the 40ms [da] would encourage its' clinical application as fewer epochs combined with the shorter stimulus duration would require shorter testing sessions than longer duration stimuli combined with more epochs.

Due to this variability in the number of epochs, implementing an automated method such as the combination of F_{SP} and bootstrap during speech-ABR recording would assist clinicians and researchers in identifying the number of epochs required for a particular individual, in addition to being confident that responses are present and that identified/detected peaks are above the background EEG noise. Applying such methods online while recording would save time in those that require fewer epochs and would increase the likelihood of response detection in those that require a larger number of epochs. Bootstrap approaches have the advantage over the F_{SP} in that they are less influenced by variability in recordings between participants; however they are more computationally complex to implement. Therefore, applying F_{SP} online during recording until a certain criterion is reached (e.g. 3.1), then applying bootstrap online after this criterion is reached would likely be more feasible. However, more work is needed to determine the appropriate F_{SP} values that correspond to 99% confidence response presence in speech-ABRs in quiet and in noise and to determine the most sensitive measure for detection of speech-ABRs.

Conclusions

This is the first study that systematically investigated the clinical feasibility of speech-ABRs as an objective audiological measure. The speech-ABR was evaluated in terms of stimulus duration, background noise, and number of epochs within the same participants. The results show that the 40ms [da] in quite and in noise is the most appropriate stimuli for the clinical implementation of the speech-ABR to evaluate speech detection and speech-ABRs in noise, based on the following:

(i) The influence of background on peak latencies and amplitudes is similar across stimuli regardless of duration, with no effect of background on the FFR in speech-ABRs to longer duration stimuli;

(ii) The lack of peak latency differences in speech-ABRs between the 3 CVs (regardless of duration) suggests that the speech-ABR may not be an appropriate tool to assess auditory discrimination of the CV stimuli used in this study.

(iii) Fewer epochs are required to record speech-ABRs with clearly identifiable peaks to the 40ms [da], this combined with the short stimulus duration leads to shorter session times.

Future Directions

Several features of the speech signal may be recorded via speech-ABRs, these include: (i) sound onset; (ii) frequency transitions; (iii) formant structure; and (iv) F_0 (Kraus & Nicol 2005; Abrams & Kraus 2015). Such features cannot be measured using current clinical click and tone burst ABRs. The speech-ABR could therefore be a valuable clinical tool in the assessment of subcortical encoding of speech in quiet and in background noise. In this study, 4 issues related to the clinical feasibility of speech-ABRs were addressed: (1) stimulus duration, (2) background (quiet versus noise), (3) CV, and (4) number of epochs. Results from this study add to existing speech-ABR literature and are a step forward towards the development of clinical protocols for speech-ABRs. More specifically, for speech-ABRs as a measure of subcortical encoding of speech and of speech-in-noise performance. However, ample work is still needed before speech-ABRs can be introduced to clinical practice. For example, prior to the clinical application of speech-ABRs as a measure of speech-in-noise performance, stimulus specific normative data on speech-ABRs in quiet versus in noise in normal hearing individuals and in clinical populations (e.g. individuals with hearing loss) are necessary, such studies should ideally include criteria for what is considered a normal change in speech-ABRs with the addition of background noise and what would indicate degradation in speech-in-noise performance. Further investigation is also needed using CVs different than those used in this study to evaluate the speech-ABRs usability as a measure of discrimination of speech sounds prior to its' clinical application for this purpose. Finally, there is a need to establish a sensitive clinically feasible measure for speech-ABR detection and confirmation of response presence (e.g. appropriate F_{SP} values that correspond to 99% confidence response presence combined with bootstrap).

ACKNOWLEDGEMENTS

We thank Dr. Timothy Wilding, Dr. Emanuele Perugia, and Dr. Frederic Marmel at the Manchester Centre for Audiology and Deafness, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre for their help in writing the MATLAB code for some of the data processing.

We also thank the Auditory Neuroscience Laboratory, Department of Communication Sciences, Northwestern University, Evanston, IL, USA for the provision of stimuli (CVs and background babble) used in this study.

We thank the Saudi Arabian Ministry of Education, King Fahad Medical City, and the Engineering and Physical Sciences Research Council (EPSRC grant No. EP/M026728/1) for funding this project.

Raw EEG data (Speech-ABRs) for this study may be accessed at (**Link to raw data to be inserted here**)

List of Supplemental Digital Content:

1. BinKhamis et al Supplemental Digital Content 1.pdf
2. BinKhamis et al Supplemental Digital Content 2.pdf

REFERENCES

Abrams, D., Kraus, N. (2015). Auditory pathway representations of speech sounds in humans. In J. Katz, M. Chasin, K. English, et al., eds. Handbook of clinical audiology. (pp. 527–544). Philadelphia: Wolters Kluwer Health.

Akhoun, I., Gallégo, S., Moulin, A., et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology*, 119, 922–933.

AlOsman, R., Giguère, C., Dajani, H. (2016). Effects of stimulus rate and noise on speech-evoked auditory brainstem responses. *Canadian Journal of Speech-Language Pathology and Audiology CJSLPA*, 40, 1–16.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., et al. (2013a). Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance. *J Speech Lang Hear Res*, 56, 31–14.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., et al. (2013b). Effects of hearing loss on the subcortical representation of speech cues. *J. Acoust. Soc. Am.*, 133, 3030–3038.

Anderson, S., Parbery-Clark, A., Yi, H., et al. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing*, 32, 750–757.

Banai, K., Hornickel, J., Skoe, E., et al. (2009). Reading and subcortical auditory function. *Cerebral Cortex*, 19, 2699–2707.

Bates, D., Mächler, M., Bolker, B., et al. (2015). Fitting linear mixed-effects models using lme4.

J. Stat. Soft., 67, 1–48.

Bellier, L., Veuillet, E., Vesson, J., et al. (2015). Speech auditory brainstem response through hearing aid stimulation. *Hearing Research*, 325, 49–54.

Burkard, R., Sims, D. (2002). A Comparison of the Effects of Broadband Masking Noise on the Auditory Brainstem Response in Young and Older Adults. *Am J Audiol*, 11, 13–10.

Chandrasekaran, B., Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, 47, 236–246.

Coffey, E., Herholz, S., Chepesiuk, A., et al. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, 7, 1–11.

Cohen, M.X. (2014). *Analyzing Neural Time Series Data*, Cambridge, Massachusetts: Massachusetts Institute of Technology. Available at: https://mitpress.mit.edu/books/analyzing-neural-time-series-data.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21.

Don, M., Elberling, C., Waring, M. (1984). Objective detection of averaged auditory brainstem responses. *Scandinavian Audiology*, 13, 219–228.

Efron, B. (1979a). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1–26.

Efron, B. (1979b). Computers and the theory of statistics: Thinking the unthinkable. *SIAM Rev.*, 21, 460–480.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *Can. J. Statistics*, 9, 139–158.

Elberling, C., Don, M. (1984). Quality estimation of averaged auditory brainstem responses. *Scandinavian Audiology*, 13, 187–197.

Galambos, R., Despland, P. (1980). The Auditory Brainstem Response (ABR) evaluates risk factors for hearing loss in the newborn. *Pediatr Res*, 14, 159–163.

Hall, J., III (2015). *eHanbook of Auditory Evoked Responses* Kindle Edition. M. Hall, ed., Pearson Educatino Inc.

Hood, L. (2015). Auditory brainstem response: Estimation of hearing sensitivity. In J. Katz, M. Chasin, K. English, et al., eds. *Handbook of Clinical Audiology*. (*pp. 249–230*). Philadelphia: Wolters Kluwer Health.

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

Hornickel, J., Skoe, E., Kraus, N. (2009a). Subcortical laterality of speech encoding. *Audiol. Neurootol*., 14, 198–207.

Hornickel, J., Skoe, E., Nicol, T., et al. (2009b). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences*, 106, 13022–13027.

Jewett, D., Romano, M., Williston, J. (1970). Human Auditory Evoked Potentials: Possible Brain

Stem Components Detected on the Scalp. *Science*, 167, 1517–1518.

Johnson, K., Nicol, T., Zecker, S., et al. (2008). Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology*, 119, 2623–2635.

Johnson, K., Nicol, T., Kraus, N. (2005). Brain stem response to speech: A biological marker of auditory processing. *Ear and Hearing*, 26, 424–434.

Kraus, N., Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences*, 28, 176–181.

Krishnan, A., Xu, Y., Gandour, J., et al. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25, 161–168.

Krizman, J., Skoe, E., Kraus, N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiol Neurotol*, 15, 332–342.

Kuznetsova, A., Brockhoff, P., Christensen, R. (2016). lmerTest: Tests in linear mixed effects models [R package lmerTest version 2.0-33].

Lenth, R. (2016). Least-squares means: the R package lsmeans. *J. Stat. Soft.*, 69, 1–33.

Li, X., Jeng, F. (2011). Noise tolerance in human frequency-following responses to voice pitch. *J. Acoust. Soc. Am.*, 129, EL21–EL26.

Liu, L., Palmer, A., Wallace, M. (2006). Phase-Locked responses to pure tones in the inferior colliculus. *J Neurophysiol*, 95, 1926–1935.

Lv, J., Simpson, D., Bell, S. (2007). Objective detection of evoked potentials using a bootstrap technique. *Medical Engineering & Physics*, 29, 191–198.

Mehraei, G., Hickox, A., Bharadwaj, H., et al. (2016). Auditory Brainstem Response Latency in Noise as a Marker of Cochlear Synaptopathy. *J. Neurosci*., 36, 3755–3764.

Nuttall, H., Moore, D., Barry, J., et al. (2015). The influence of cochlear spectral processing on the timing and amplitude of the speech-evoked auditory brain stem response. *J Neurophysiol*, 113, 3683–3691.

Parbery-Clark, A., Marmel, F., Bair, J., et al. (2011). What subcortical-cortical relationships tell us about processing speech in noise. *Eur J Neurosci*, 33, 549–557.

Prévost, F., Laroche, M., Marcoux, A., et al. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology*, 124, 52–60.

R Core Team (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Available at: https://www.R-project.org/.

Russo, N., Nicol, T., Musacchia, G., et al. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, 115, 2021–2030.

Russo, N., Skoe, E., Trommer, B., et al. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clinical Neurophysiology*, 119, 1720–1731.

Skoe, E., Kraus, N. (2010). Auditory brain stem response to complex sounds: A tutorial. *Ear and Hearing*, 31, 302–324.

Skoe, E., Krizman, J., Anderson, S., et al. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cereb. Cortex*, 25, 1415–1426.

Skoe, E., Nicol, T., Kraus, N. (2011). Cross-phaseogram: Objective neural index of speech sound differentiation. *Journal of Neuroscience Methods*, 196, 308–317.

Smalt, C., Krishnan, A., Bidelman, G., et al. (2012). Distortion products and their influence on representation of pitch-relevant information in the human brainstem for unresolved harmonic complex tones. *Hearing Research*, 292, 26–34.

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

Song, J., Skoe, E., Banai, K., et al. (2011b). Perception of speech in noise: Neural correlates. *Journal of Cognitive Neuroscience*, 23, 2268–2279.

Swaminathan, J., Krishnan, A., Gandour, J.T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport*, 19, 1163–1167.

Van Campen, L., Sammeth, C., Hall, J. III, et al. (1992). Comparison of Etymotic insert and TDH supra-aural earphones in auditory brainstem response measurement. *J am acad audiol*, 3, 315–323.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. arXiv:1308.5499. Available at: [http://arxiv.org/pdf/1308.5499.pdf].

Figure Legends

Fig. 1. Speech-ABR with pre-stimulus baseline to the 40ms [da] from one participant (12000 epochs) after bootstrap showing all peaks above/below the 95% confidence lines.

Fig. 2. Grand average speech-ABRs with pre-stimulus baseline in quiet and in noise to the: (A) 40ms [da] and (B) 170ms [da]; showing longer peak latencies and smaller peak amplitudes in noise compared to in quiet across the 2 [da] durations. Shade in all panels represents 1 SE.

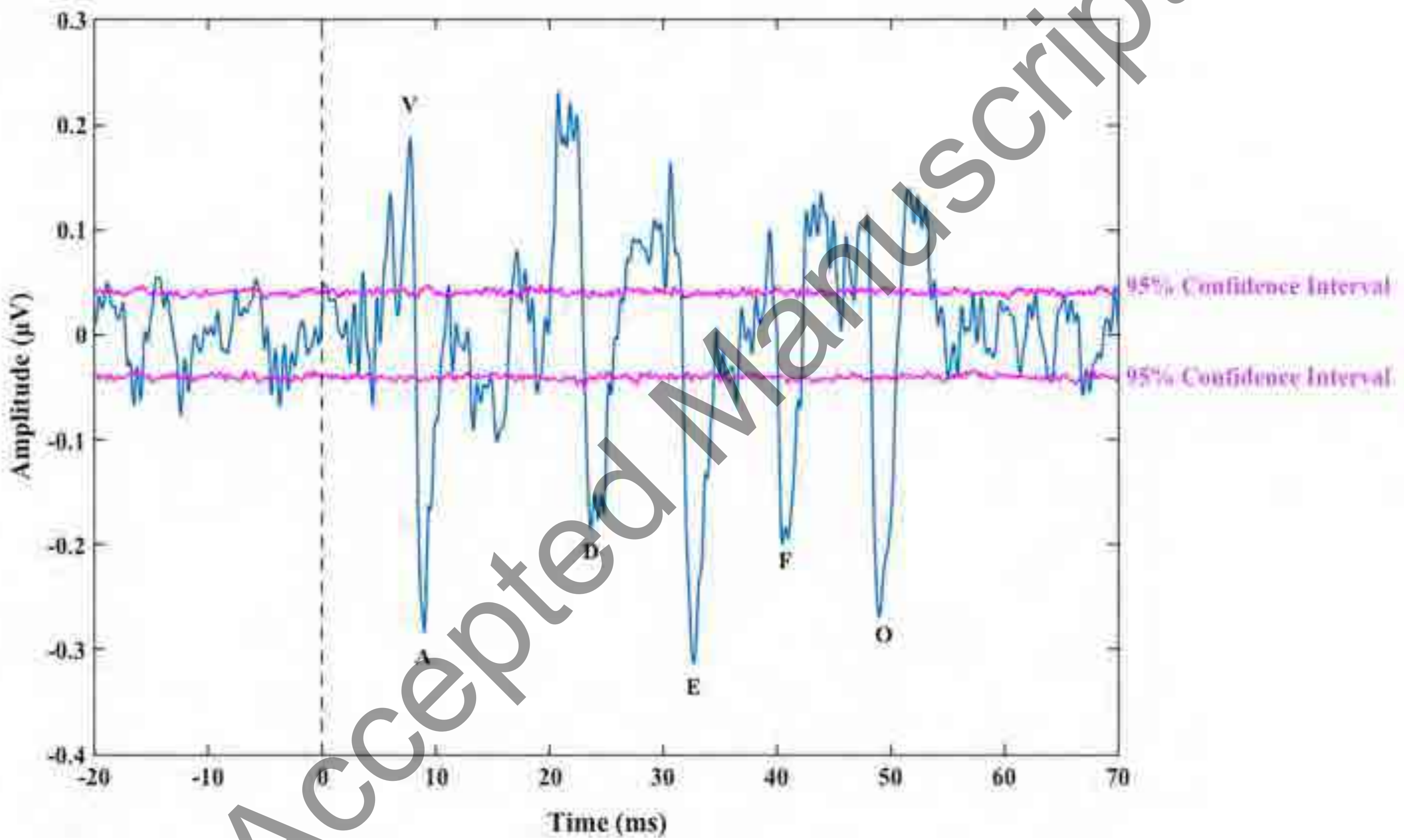
Fig. 3. Grand average speech-ABRs with pre-stimulus baseline in quiet and in noise to the: (A) 50ms [ba], (B) 50ms [da], and (C) 50ms [ga]; showing longer peak latencies and smaller peak amplitudes in noise compared to in quiet across the 3 CVs. Shade in all panels represents 1 SE.

Fig. 4. Grand average speech-ABRs in quiet and in noise to the 170ms [da]: (A) time domain waveforms with pre-stimulus baseline, showing greater effect of noise on onset and transition peaks (0 – 70ms) than on the later FFR period (70-190ms), (B) FFR degree of phase locking to F_0, showing a non-significant trend for higher degree of phase locking to F_0 in noise compared to quiet. Shade in all panels represents 1 SE. And (C) spectrum (FFT) of the onset and transition period (0-70ms) showing the greater effect of noise on F_0 in the first 70ms or the response.

Fig. 5. Grand average speech-ABRs with pre-stimulus baseline in quiet to the: (A) 50ms [ba] [da] [ga] and (B) 170ms [ba] [da] [ga]; showing no differences in peak latencies and amplitudes between the 2 stimulus durations, and no differences in peak latencies between the 3 CVs across

the 2 stimulus durations. While responses to 170ms [ba] appear to have longer peak latencies, this was not significant. Shade in all panels represents 1 SE.

Figure_1



0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Amplitude (μV)
-20
-10
0
10
20
30
40
50
60
70
Time (ms)
V
A
D
E
F
O
95% Confidence Interval
95% Confidence Interval

Figure_2

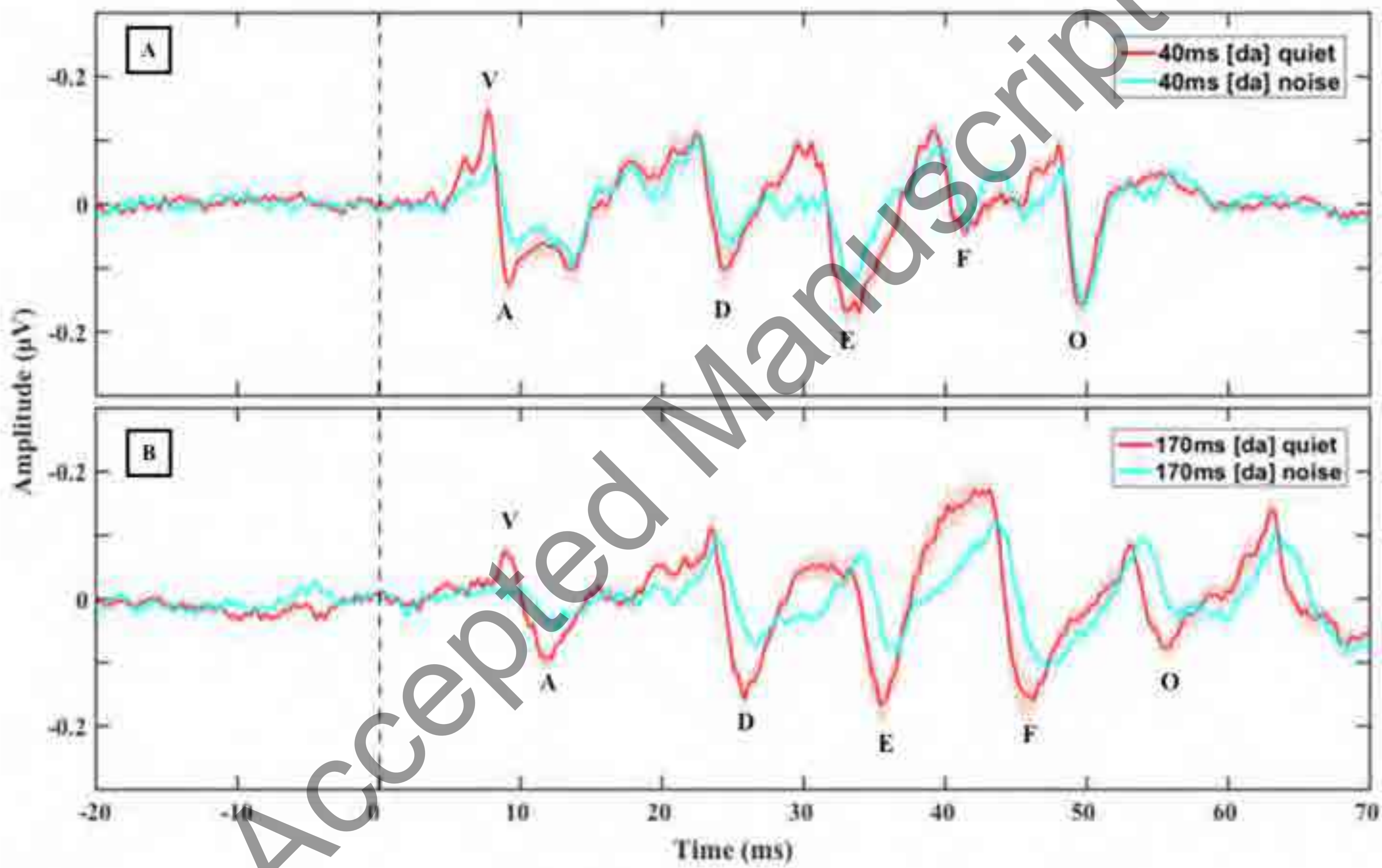
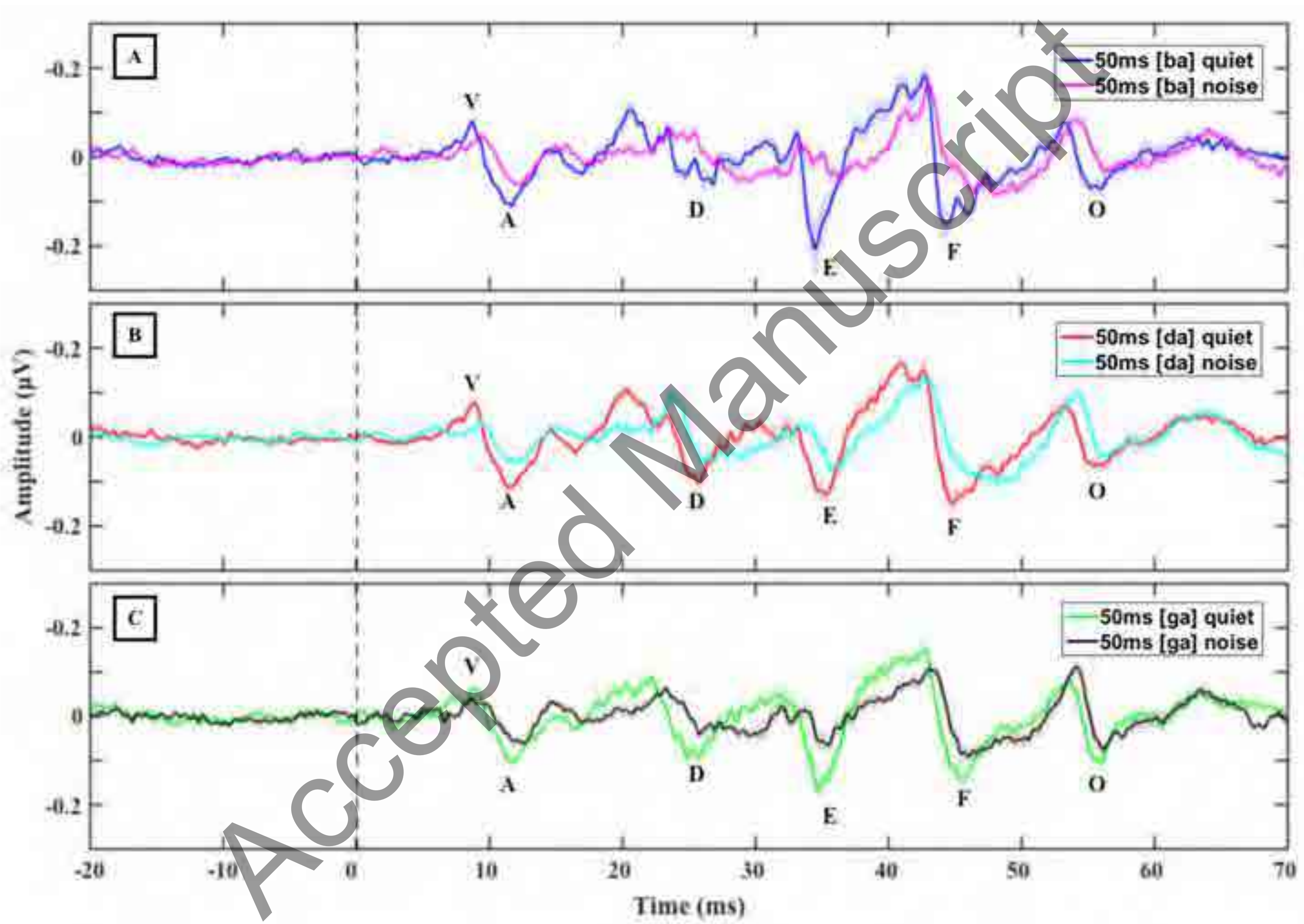

A
40ms [da] quiet
40ms [da] noise
V
A
D
E
F
O
B
170ms [da] quiet
170ms [da] noise
V
A
D
E
F
O
-0.2
0
-0.2
Amplitude (µV)
-20
-10
0
10
20
30
40
50
60
70
Time (ms)

Figure_3



A
B
C
50ms [ba] quiet
50ms [ba] noise
50ms [da] quiet
50ms [da] noise
50ms [ga] quiet
50ms [ga] noise
V
A
D
E
F
O
Amplitude (µV)
Time (ms)
-0.2
0
0.2
-20
-10
0
10
20
30
40
50
60
70

Figure_4

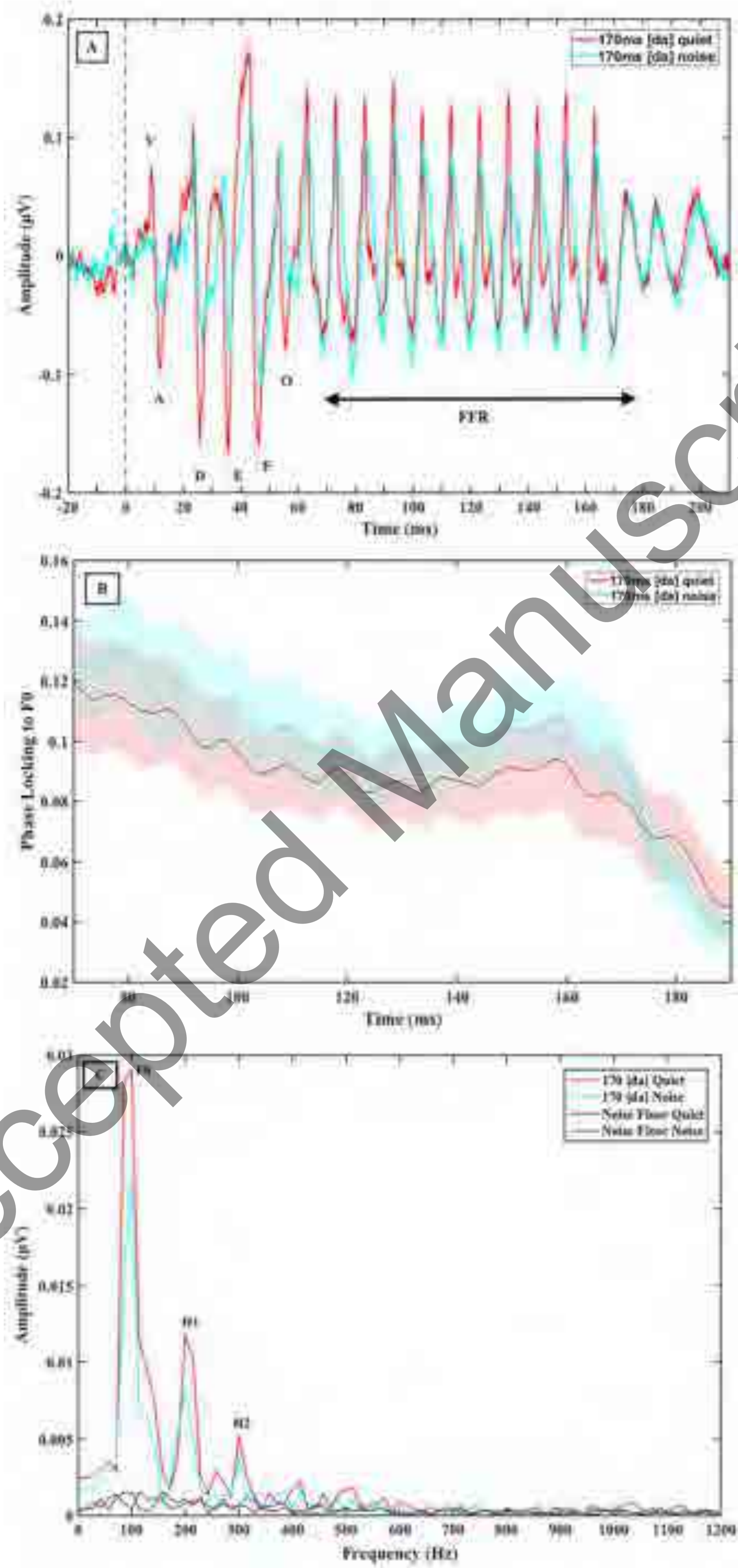
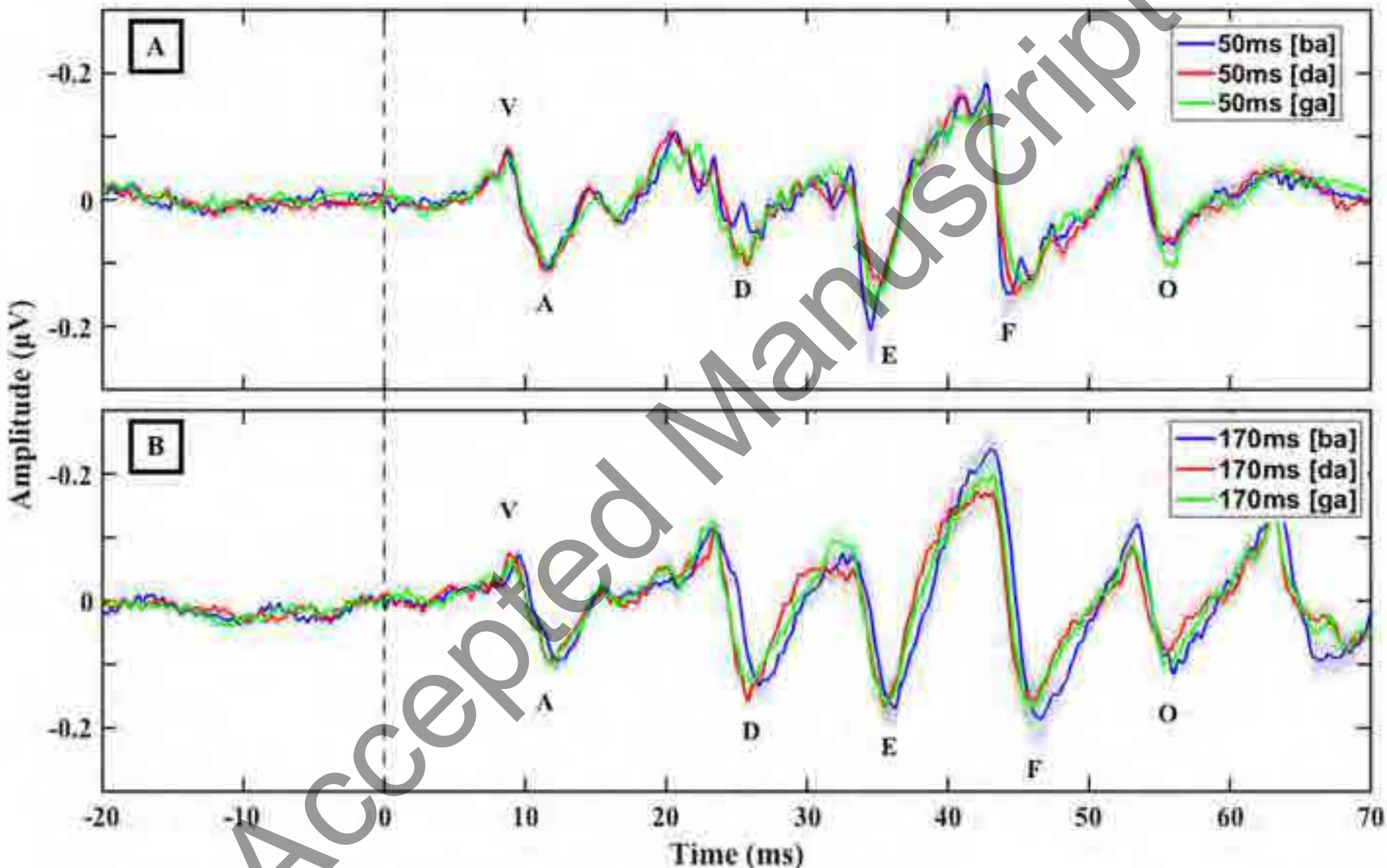

A
170ms [da] quiet
170ms [da] noise
V
A
D
E
F
O
FFR
Amplitude (µV)
Time (ms)
B
Phase Locking to F0
Time (ms)
C
F0
H1
H2
170 [da] Quiet
170 [da] Noise
Noise Floor Quiet
Noise Floor Noise
Amplitude (µV)
Frequency (Hz)

Figure_5



A
B
V
A
D
E
F
O
50ms [ba]
50ms [da]
50ms [ga]
170ms [ba]
170ms [da]
170ms [ga]
Amplitude (µV)
Time (ms)
-0.2
0
0.2
-20
-10
0
10
20
30
40
50
60
70

Table 1. Speech-ABR Recording Parameters including presentation rate (in stimulus per second), epoch length (recording time window including inter stimulus interval, in milliseconds), sound intensity, high-pass and low-pass filter cut-off frequency (Hz), and amplification (gain).

	40ms [da]	50ms CVs	170ms CVs
Rate	11.1	9.1	4.35
Epoch Length	90	110	230
Intensity (dB A)	80	80	80
High-pass filter*	100	70	
Low-pass filter	3000		
Gain	10000		

* High-pass filter settings were set based on values used in the 40ms [da] literature (e.g. Anderson et al. 2013a; Skoe et al. 2015) and in the 170ms [ba] [da] [ga] literature (e.g. Johnson et al. 2008; Hornickel et al. 2009b; Anderson et al. 2011; Parbery-Clark et al. 2011)

Supplemental Digital Content 1

Speech Auditory Brainstem Responses: Effects of Background, Stimulus Duration, Consonant-Vowel, and Number of Epochs

Section 1: Characteristics of CV Stimuli

Table 1. Formant Frequency Components ($F_0 – F_6$, in Hz) of CV Stimuli used to Record Speech-ABRs.

↑ = *frequency rises during vowel formant transition.*

↓ = *frequency falls during vowel formant transition.*

	40ms	50ms			170ms		
	[da]	**[ba]**	**[da]**	**[ga]**	**[ba]**	**[da]**	**[ga]**
F_0	103 ↑ 125	100			100		
F_1	220 ↑ 720	400 ↑ 720			400 ↑ 720		
F_2	1700 ↓ 1240	900 ↑ 1240	1700 ↓ 1240	2480 ↓ 1240	900 ↑ 1240	1700 ↓ 1240	2480 ↓ 1240
					Frequency transition during the first 50ms		
F_3	2580 ↓ 2500	2580 ↓ 2500			2580 ↓ 2500		
F_4	3600	3300			3300		
F_5	4500	3750			3750		
F_6	NA	4900			4900		

40ms [da] from (Banai et al. 2009)

170ms [ba] [da] [ga] from (Hornickel et al. 2009)

Time Domain Waveforms of Stimuli

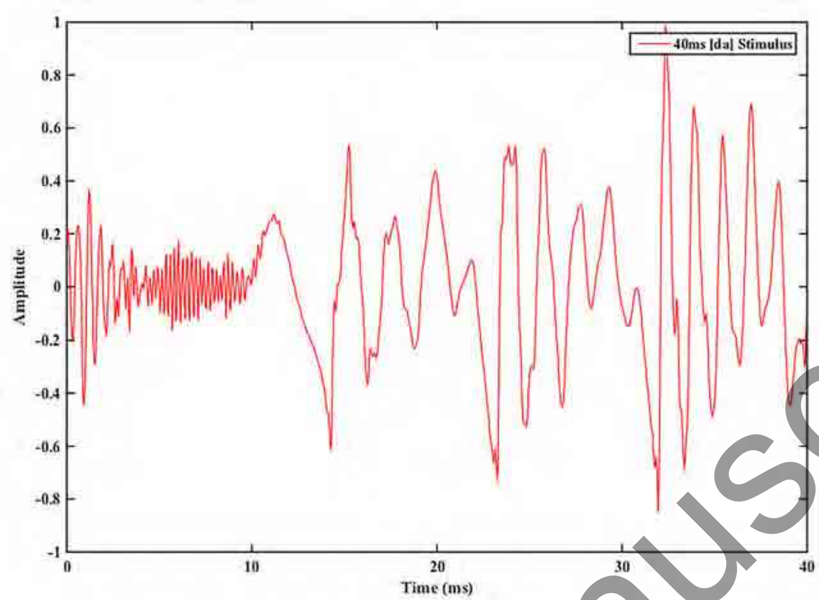


Fig. 1. Time domain waveform of a single polarity 40ms [da] stimulus

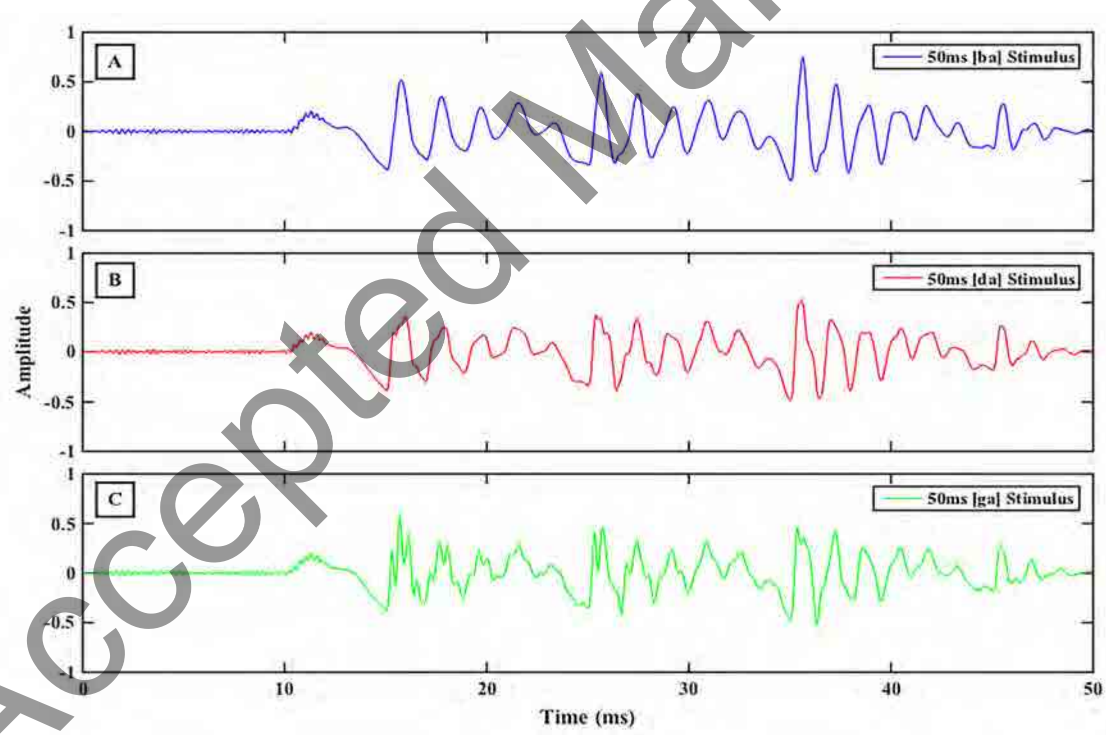


Fig. 2. Time domain waveforms of a single polarity: (A) 50ms [ba] stimulus, (B) 50ms [da] stimulus, (C) 50ms [ga] stimulus

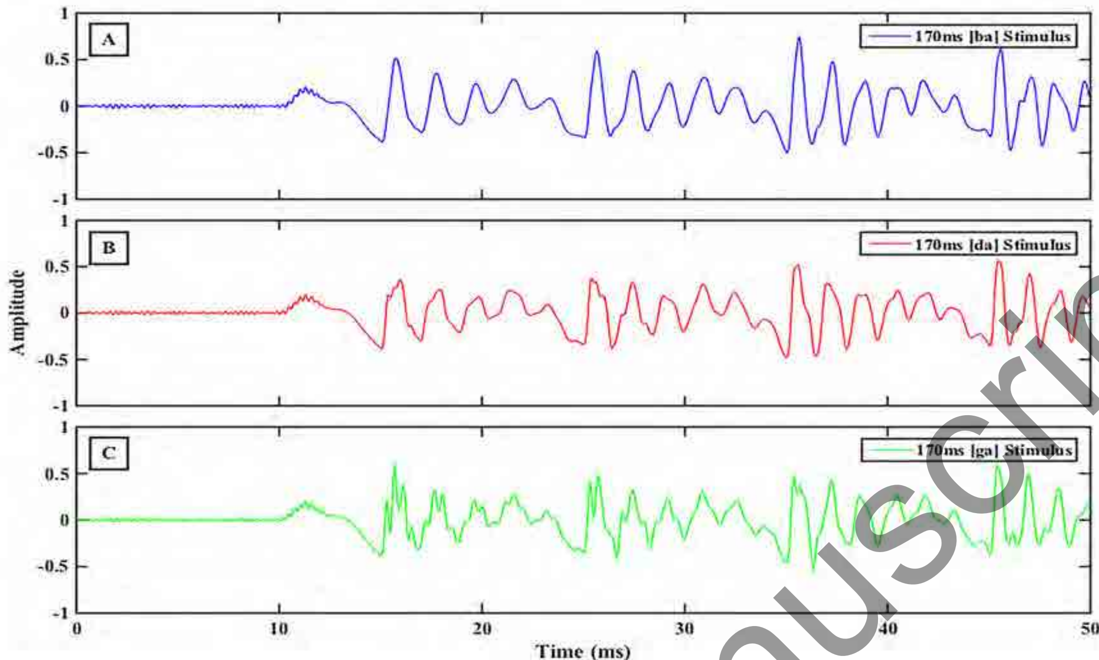


Fig. 3. Time domain waveforms of the transition period (first 50ms) of a single polarity: (A) 170ms [ba] stimulus, (B) 170ms [da] stimulus, (C) 170ms [ga] stimulus

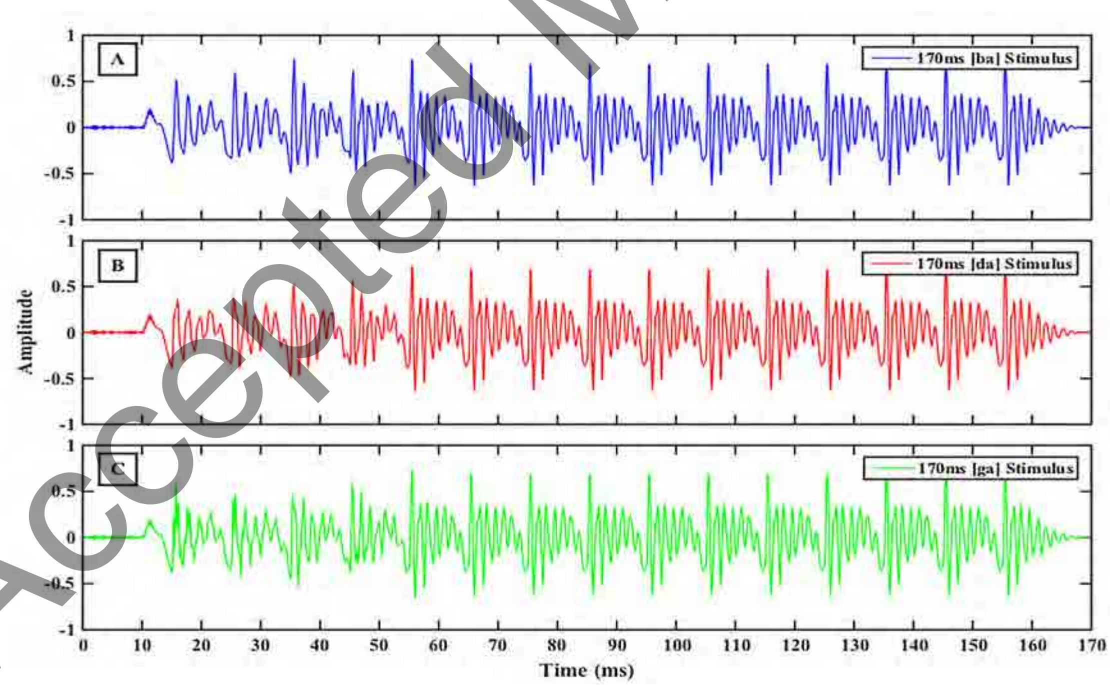


Fig. 4. Time domain waveforms of a single polarity: (A) 170ms [ba] stimulus, (B) 170ms [da] stimulus, (C) 170ms [ga] stimulus

Spectrum of Stimuli

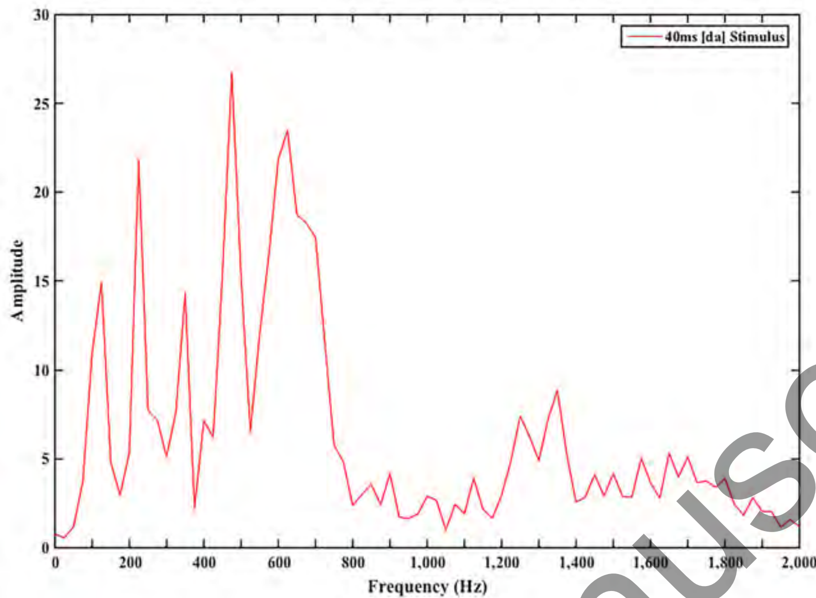


Fig. 5. Spectrum (FFT of the full stimulus) of a single polarity 40ms [da] stimulus

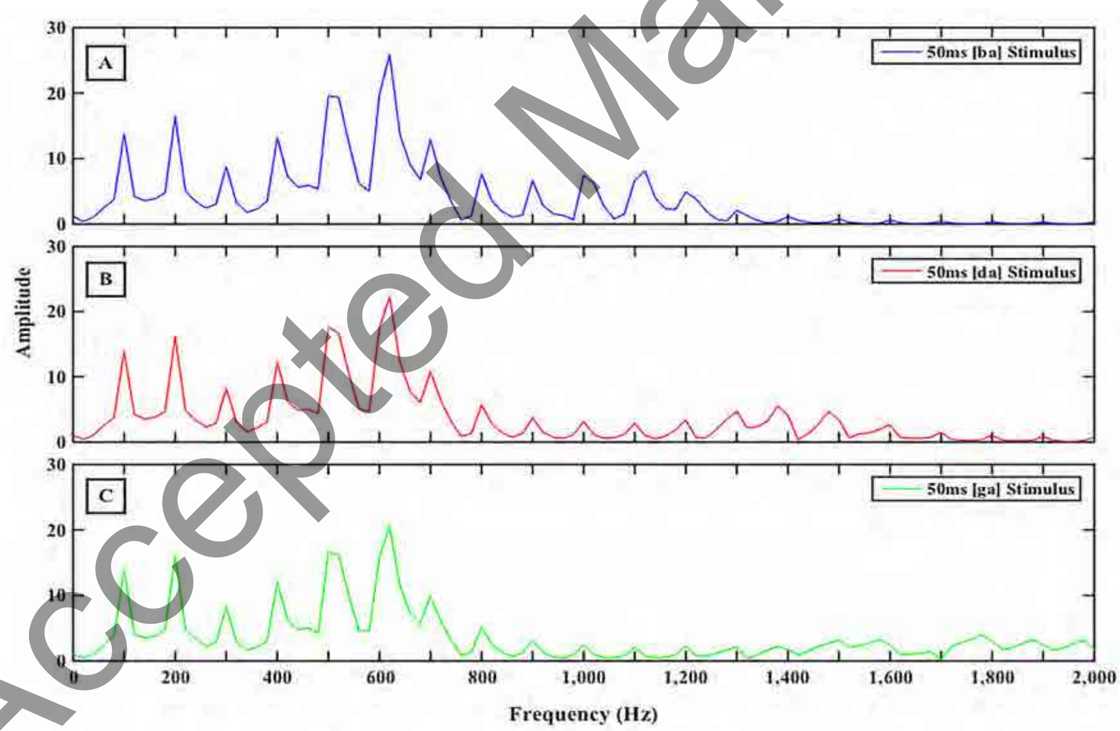


Fig. 6. Spectrum (FFT of the full stimulus) of a single polarity: (A) 50ms [ba] stimulus, (B) 50ms [da] stimulus, (C) 50ms [ga] stimulus

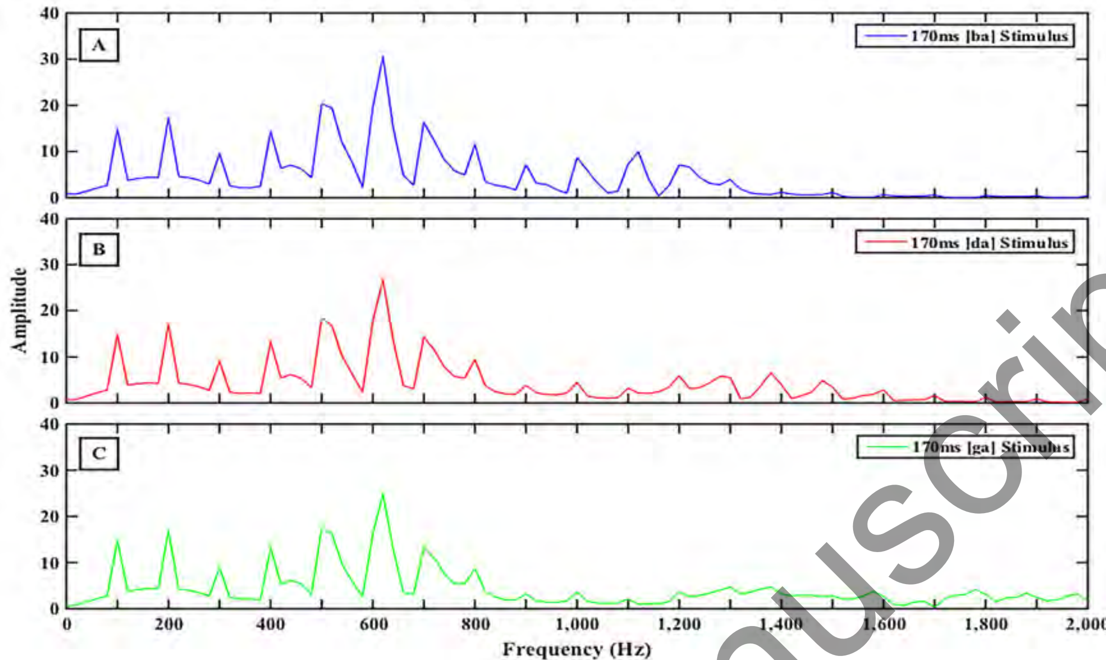


Fig. 7. Spectrum (FFT of the first 50ms of the stimulus) of the transition period (first 50ms) of a single polarity: (A) 170ms [ba] stimulus, (B) 170ms [da] stimulus, (C) 170ms [ga] stimulus

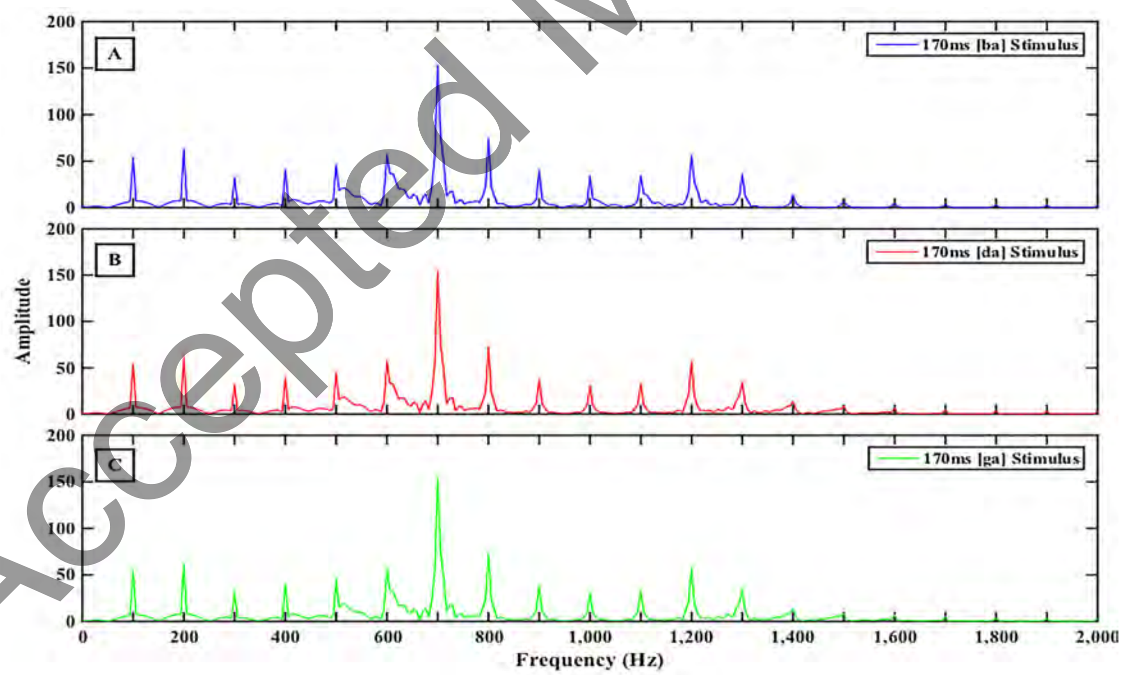


Fig. 8. Spectrum (FFT of the full stimulus) of a single polarity: (A) 170ms [ba] stimulus, (B) 170ms [da] stimulus, (C) 170ms [ga] stimulus

Section 2: Recording Time Per Stimulus

Table 2. Mean, SD, and Range of Recording Times (in minutes) required for completing 4 Speech-ABR blocks (i.e. 12,000 epochs) per stimulus, across durations (40ms, 50ms, 170ms), and in each background (quiet and noise).
Shaded cells indicate that stimulus was not tested in noise.

		Quiet			Noise		
		Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
40ms	**[da]**	23.17	5.22	20 – 31	21.75	3.19	19 – 25
50ms	**[ba]**	27.08	3.94	24 – 36	30.17	7.52	25 – 49
	[da]	27.75	6.45	24 – 45	28.17	5.64	23 – 45
	[ga]	29.17	5.75	25 – 40	30.33	7.45	25 – 47
170ms	**[ba]**	56.08	15.92	48 – 106			
	[da]	54.50	8.92	48 – 81	56.00	7.48	49 – 71
	[ga]	53.92	7.04	49 – 73			

Section 3: Filtering Speech-ABRs to Emphasize Peak Latency Differences Between [ba], [da], and [ga]

Johnson et al. 2008 and Hornickel et al. 2009 reported first band-pass filtering speech-ABRs to each stimulus polarity from 70 – 2000 Hz, then adding speech-ABRs to the 2 polarities. Following filtering and adding speech-ABRs, an additional high-pass filter of 300 Hz was applied to the added speech-ABRs.

In speech-ABRs that were recorded in this study, applying the additional high-pass filter to the added responses resulted in a drastic decrease in speech-ABR amplitudes with no clearly defined peaks (Figs.9, 10, 11, 12). A spectrum of speech-ABR onset and transition periods to these 3 stimuli shows that speech-ABRs from this study have little to no spectral peaks above 300 Hz (Fig. 13), which explains why responses were obliterated when high-pass filtered at 300 Hz.

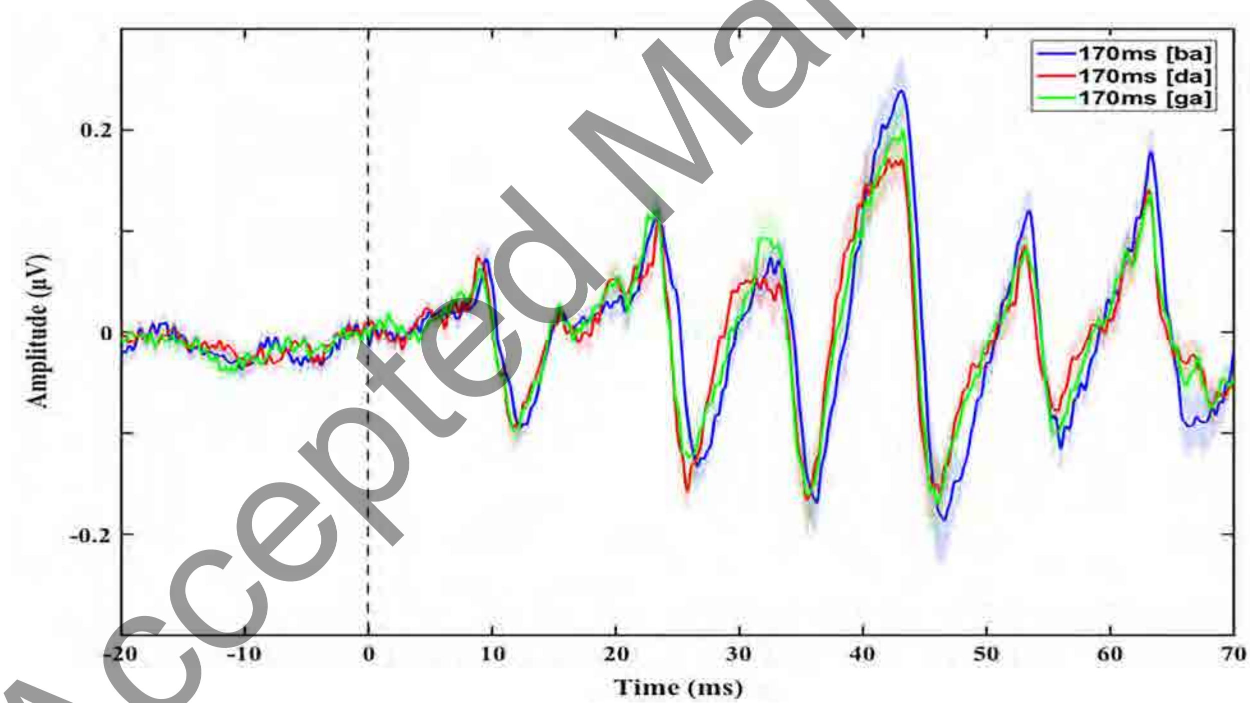


Fig. 9. Grand average speech-ABRs with pre-stimulus baseline (onset and transition period: 0 – 70ms) in quiet to the 170ms [ba] [da] [ga] overlaid, band-pass filtered 70 – 2000 Hz. Shade represents 1 SE.

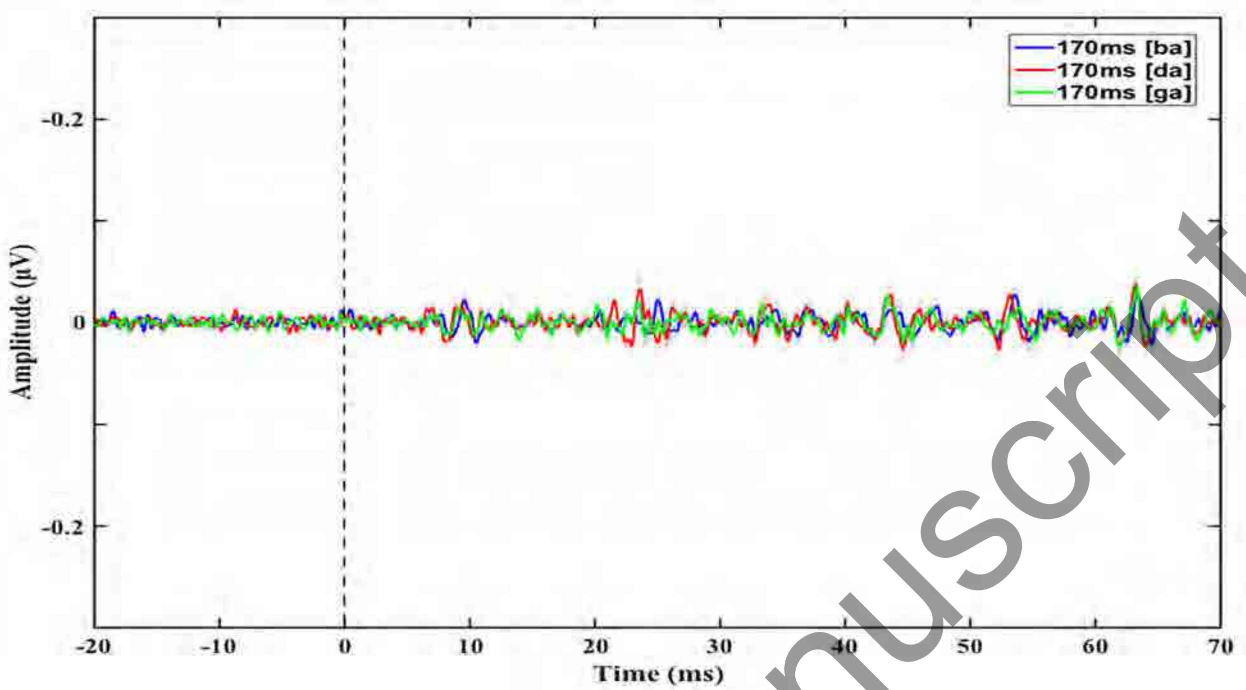


Fig. 10. Grand average speech-ABRs with pre-stimulus baseline (onset and transition period: 0 – 70ms) in quiet to the 170ms [ba] [da] [ga] overlaid, with additional high-pass filter (300 Hz) applied, showing the drastic decrease in amplitudes and overall absence of responses. Shade represents 1 SE

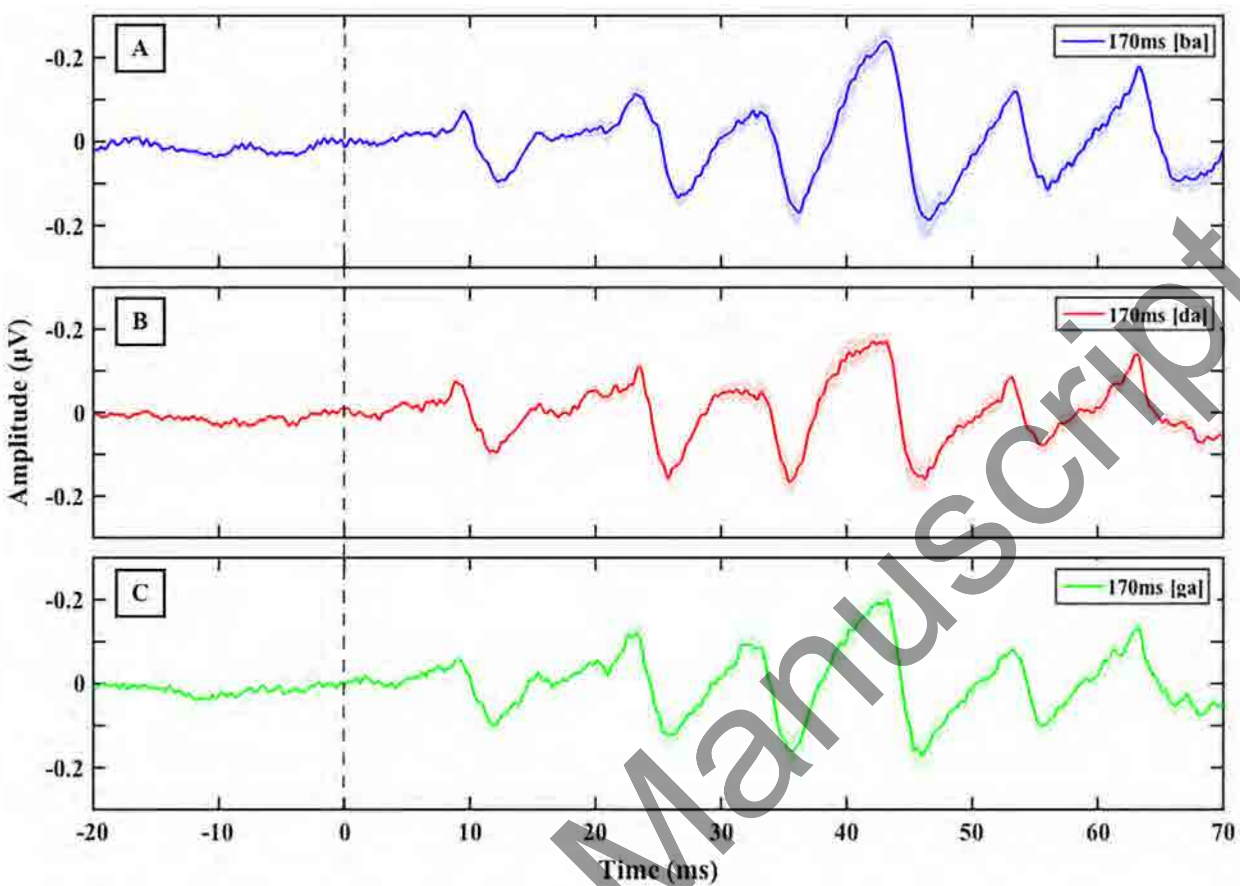


Fig. 11. Grand average speech-ABRs with pre-stimulus baseline (onset and transition period: 0 – 70ms) in quiet to the: (A) 170ms [ba], (B) 170ms [da], (C) 170ms [ga] plotted separately, band-pass filtered 70 – 2000 Hz. Shade in all panels represents 1 SE.

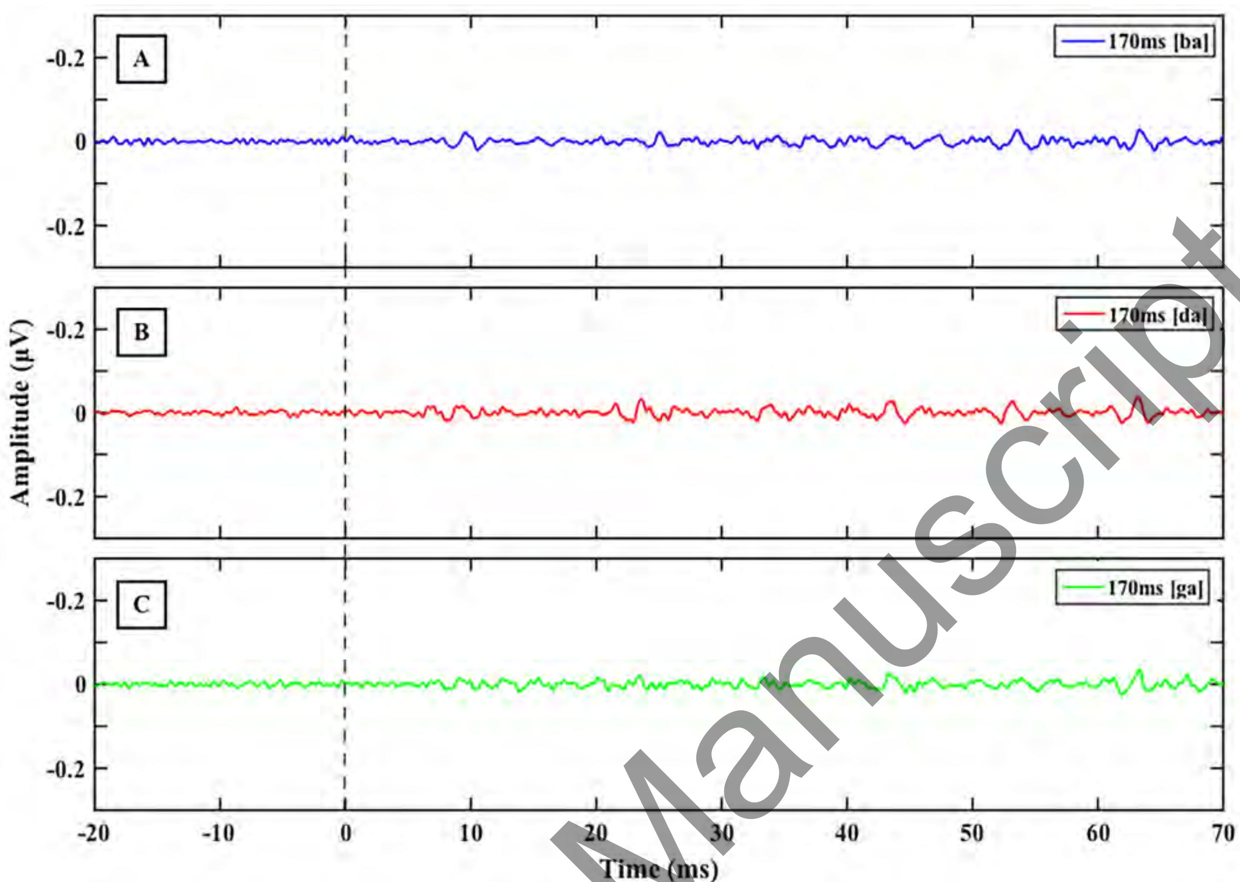


Fig. 12. Grand average speech-ABRs with pre-stimulus baseline (onset and transition period: 0 – 70ms) in quiet to the: (A) 170ms [ba], (B) 170ms [da], (C) 170ms [ga] plotted separately, with additional high-pass filter (300 Hz) applied. Shade in all panels represents 1 SE.

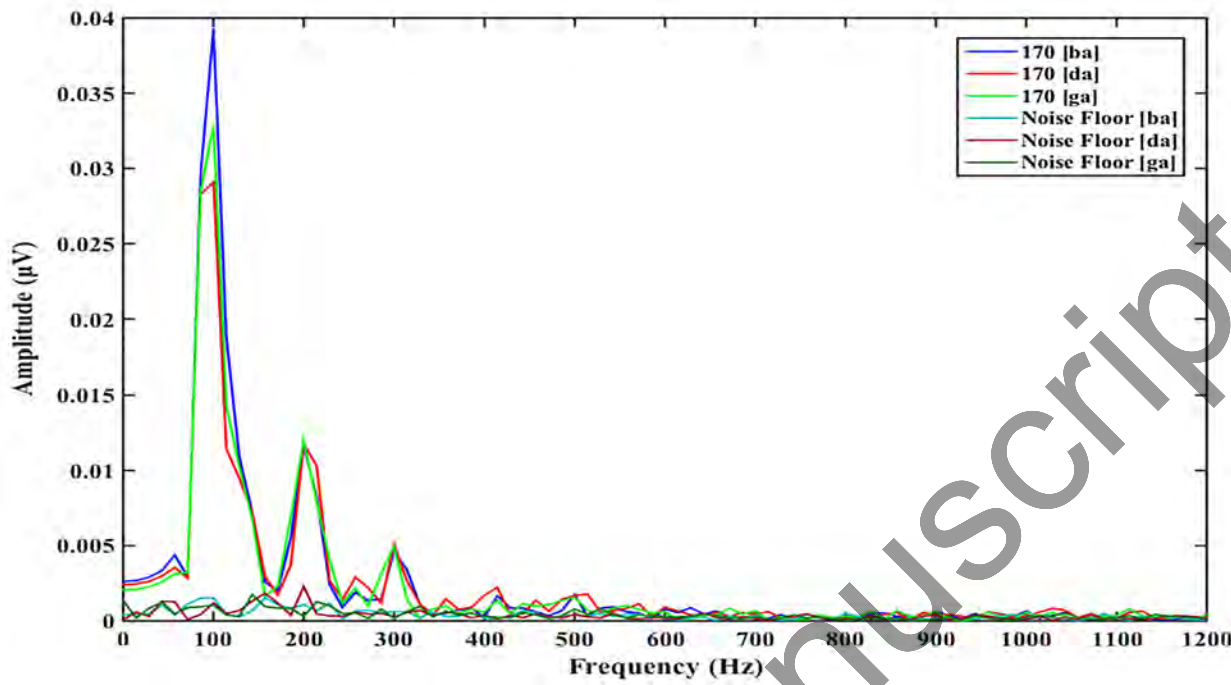


Fig. 13. Spectrum of grand average speech-ABRs band-pass filtered 70 – 2000 Hz (FFT of onset and transition period: 0 – 70ms) in quiet to the 170ms [ba], [da], and [ga] showing little to no spectral peaks above 300 Hz.

Section 4: Why Speech-ABRs Contained No Spectral Peaks Above 300 Hz

In order to best predict the expected spectra of the speech-ABRs, half-wave rectifying the acoustic signals of the 2 stimulus polarities then processing their waveforms through the same analyses as the speech-ABR raw data provides a prediction of the spectral characteristics of the speech-ABR in idealized circumstances (i.e. if the auditory system encodes the acoustic waveform with absolute accuracy). Therefore, for the acoustic stimulus spectra to be comparable to the speech-ABR spectra, the 170ms [ba] [da] [ga] acoustic stimuli were processed similarly to the speech-ABRs for comparison. The following steps were conducted:

1. For each stimulus (170ms [ba], 170ms [da], and 170ms [ga]), each stimulus polarity was half-wave rectified.
2. The half-wave rectified 2 polarities of each stimulus were added (as speech-ABRs to the 2 stimulus polarities were added).
3. FFTs were performed on the transition period (first 50ms) of the added half-wave rectified stimuli.

The resulting spectra of the half-wave rectified added stimuli (Fig. 14) are similar to the speech-ABR spectra, i.e. they contain 3 peaks at 100 Hz, 200 Hz, and 300 Hz and no clear spectral peaks above 300 Hz. It would therefore not be expected for the speech-ABRs to these stimuli to contain any spectral peaks above 300 Hz.

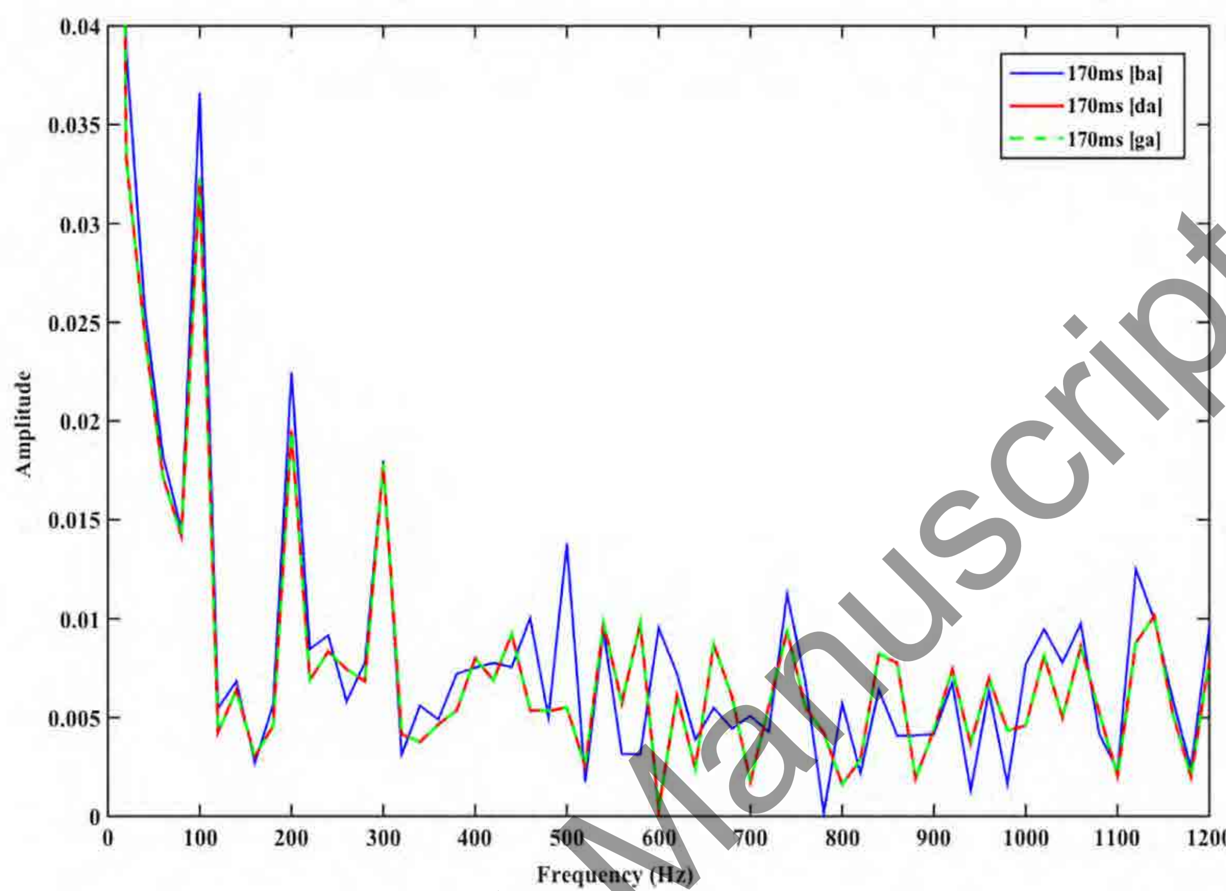


Fig. 14. Spectrum (FFT of the first 50ms) of the transition period of the half-wave rectified and added 170ms [ba] [da] and [ga] stimuli, showing 3 peaks at 100 Hz, 200 Hz, and 300 Hz with little to no spectral content above 300 Hz.

References:

Banai, K., Hornickel, J., Skoe, E., et al. (2009). Reading and subcortical auditory function. *Cerebral Cortex*, 19, 2699–2707.

Hornickel, J., Skoe, E., Nicol, T., et al. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences*, 106, 13022–13027.

Johnson, K., Nicol, T., Zecker, S., et al. (2008). Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology*, 119, 2623–2635.

Supplemental Digital Content 2

Speech Auditory Brainstem Responses: Effects of Background, Stimulus Duration, Consonant-Vowel, and Number of Epochs

Section 1: Detection of Speech-ABR Peaks

Table 1. Peaks that were missing in each participant's speech-ABRs in quiet (Q) and in noise (N) to all stimuli.
Shaded cells indicate no peaks were missing

	40ms [da]		50ms stimuli						170ms stimuli			
			[ba]		[da]		[ga]		[ba]	[da]		[ga]
Participant	**Q**	**N**	**Q**	**N**	**Q**	**N**	**Q**	**N**	**Q**	**Q**	**N**	**Q**
1		D, F										
2				O							V	
3		D		V, O			V	V, A			V	
4				V				V				
5				E				V			V, A	
6		F									V	
7			O	O		V		D				
8												
9	F	F										
10		F				V		V	V	V	V	V
11	F						D	D				
12			O	O								

Section 2: Speech-ABR Mean (SD) Peak Latencies and Amplitudes

Table 2. Mean and SD speech-ABR peak latency values (corrected for insert tube length) in quiet and in noise to the three [da] durations.

	40ms [da]				50ms [da]				170ms [da]			
	Quiet		Noise		Quiet		Noise		Quiet		Noise	
Peak	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
V	7.15	0.26	7.73	0.48	8.28	0.43	8.81	0.30	8.47	0.41	9.15	0.31
A	8.31	0.77	8.84	0.52	10.65	0.74	11.08	0.61	11.26	0.67	11.85	0.86
D	23.46	0.94	24.10	1.22	24.67	0.56	26.18	1.65	25.68	1.64	26.33	1.47
E	32.23	0.77	32.34	0.73	33.80	1.53	35.23	1.67	35.36	1.48	35.97	1.44
F	41.00	1.50	41.87	2.14	44.44	0.86	46.29	1.76	45.65	1.75	46.53	1.67
O	48.68	0.38	48.96	0.45	55.44	1.79	56.45	1.37	55.78	1.68	57.39	2.15

Table 3. Mean and SD speech-ABRs peak latency values (corrected for insert tube length) to the 50ms [ba] and [ga] in quiet and in noise, and to the 170ms [ba] and [ga] in quiet.

	50ms [ba]				170ms [ba]		50ms [ga]				170ms [ga]	
	Quiet		Noise		Quiet		Quiet		Noise		Quiet	
Peak	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
V	8.20	0.46	8.91	0.46	8.80	0.42	8.39	0.49	8.94	0.70	8.68	0.44
A	10.42	0.75	11.16	0.59	11.55	0.81	10.96	0.65	11.39	0.74	11.00	0.81
D	25.41	1.91	26.96	1.91	25.91	0.73	24.26	1.59	26.03	1.43	25.15	1.02
E	34.77	3.25	36.50	4.82	35.32	1.24	33.96	0.90	34.45	1.14	35.17	1.70
F	44.84	3.33	46.47	3.63	46.10	0.95	44.54	0.76	45.19	0.92	45.22	0.75
O	55.18	1.40	57.17	1.78	55.43	1.17	55.40	1.73	55.84	1.62	55.45	1.52

Table 4. Mean and SD speech-ABR peak amplitude values in quiet and in noise to the three [da] durations.

	40ms [da]				50ms [da]				170ms [da]			
	Quiet		Noise		Quiet		Noise		Quiet		Noise	
Peak	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
VA	0.32	0.11	0.18	0.06	0.25	0.09	0.14	0.05	0.23	0.09	0.11	0.05
D	0.31	0.15	0.20	0.18	0.28	0.10	0.22	0.09	0.34	0.12	0.23	0.07
E	0.34	0.11	0.21	0.06	0.28	0.07	0.19	0.07	0.36	0.10	0.21	0.07
F	0.23	0.10	0.16	0.13	0.42	0.10	0.34	0.09	0.47	0.16	0.30	0.06
O	0.32	0.09	0.27	0.09	0.24	0.10	0.21	0.08	0.26	0.07	0.23	0.09

Table 5. Mean and SD speech-ABR peak amplitude values to the 50ms [ba] and [ga] in quiet and in noise, and to the 170ms [ba] and [ga] in quiet.

	50ms [ba]				170ms [ba]		50ms [ga]				170ms [ga]	
	Quiet		Noise		Quiet		Quiet		Noise		Quiet	
Peak	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
VA	0.24	0.07	0.14	0.04	0.23	0.08	0.24	0.07	0.14	0.06	0.22	0.07
D	0.25	0.09	0.18	0.06	0.33	0.10	0.27	0.13	0.16	0.09	0.36	0.13
E	0.36	0.25	0.17	0.10	0.38	0.09	0.34	0.11	0.20	0.08	0.38	0.10
F	0.44	0.14	0.32	0.11	0.55	0.22	0.41	0.10	0.31	0.09	0.49	0.17
O	0.24	0.10	0.17	0.14	0.30	0.10	0.27	0.07	0.31	0.16	0.27	0.09

Section 3: Effects of Background on Speech-ABRs – Post Hoc Pairwise Comparison Results

Table 6. Post hoc pairwise comparisons of speech-ABR peak latencies comparing the two backgrounds: quiet (Q) versus noise (N) per stimulus duration, showing differences in peak latencies in quiet versus in noise (Q minus N), standard error (SE), degrees of freedom (df), *t* ratio, and bonferroni corrected *p* values.

Significant *p* values are shown in **blue**

	Stimulus duration: 40ms					Stimulus duration: 50ms					Stimulus duration: 170ms				
Peak	**Q – N (ms)**	**SE**	**df**	***t* ratio**	***p***	**Q – N (ms)**	**SE**	**df**	***t* ratio**	***p***	**Q – N (ms)**	**SE**	**df**	***t* ratio**	***p***
V	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**
A	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**
D	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**
E	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**
F	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**
O	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**	-0.91	0.10	796.07	-9.42	**0.0054**

Table 7. Post hoc pairwise comparisons of Speech-ABR peak amplitudes comparing the two backgrounds: quiet (Q) versus noise (N) per stimulus duration, showing differences in peak amplitudes in quiet versus in noise (Q minus N), standard error (SE), degrees of freedom (df), *t* ratio, and bonferroni corrected *p* values.

Significant *p* values are shown in **blue**

	Stimulus duration: 40ms					**Stimulus duration: 50ms**					**Stimulus duration: 170ms**				
Peak	**Q – N (µV)**	**SE**	**df**	***t* ratio**	***p***	**Q – N (µV)**	**SE**	**df**	***t* ratio**	***p***	**Q – N (µV)**	**SE**	**df**	***t* ratio**	***p***
VA	0.11	0.02	687	5.06	**0.0045**	0.10	0.02	687	5.73	**0.0045**	0.14	0.02	687	7.21	**0.0045**
D	0.09	0.02	687	4.19	**0.0045**	0.08	0.02	687	4.62	**0.0045**	0.12	0.02	687	6.24	**0.0045**
E	0.14	0.02	687	6.47	**0.0045**	0.13	0.02	687	7.54	**0.0045**	0.17	0.02	687	8.78	**0.0045**
F	0.12	0.02	687	5.68	**0.0045**	0.11	0.02	687	6.53	**0.0045**	0.15	0.02	687	7.91	**0.0045**
O	0.03	0.02	687	1.34	0.1823	0.02	0.02	687	0.98	0.3263	0.06	0.02	687	3.07	0.1035

Section 4: Bootstrap Results and Examples

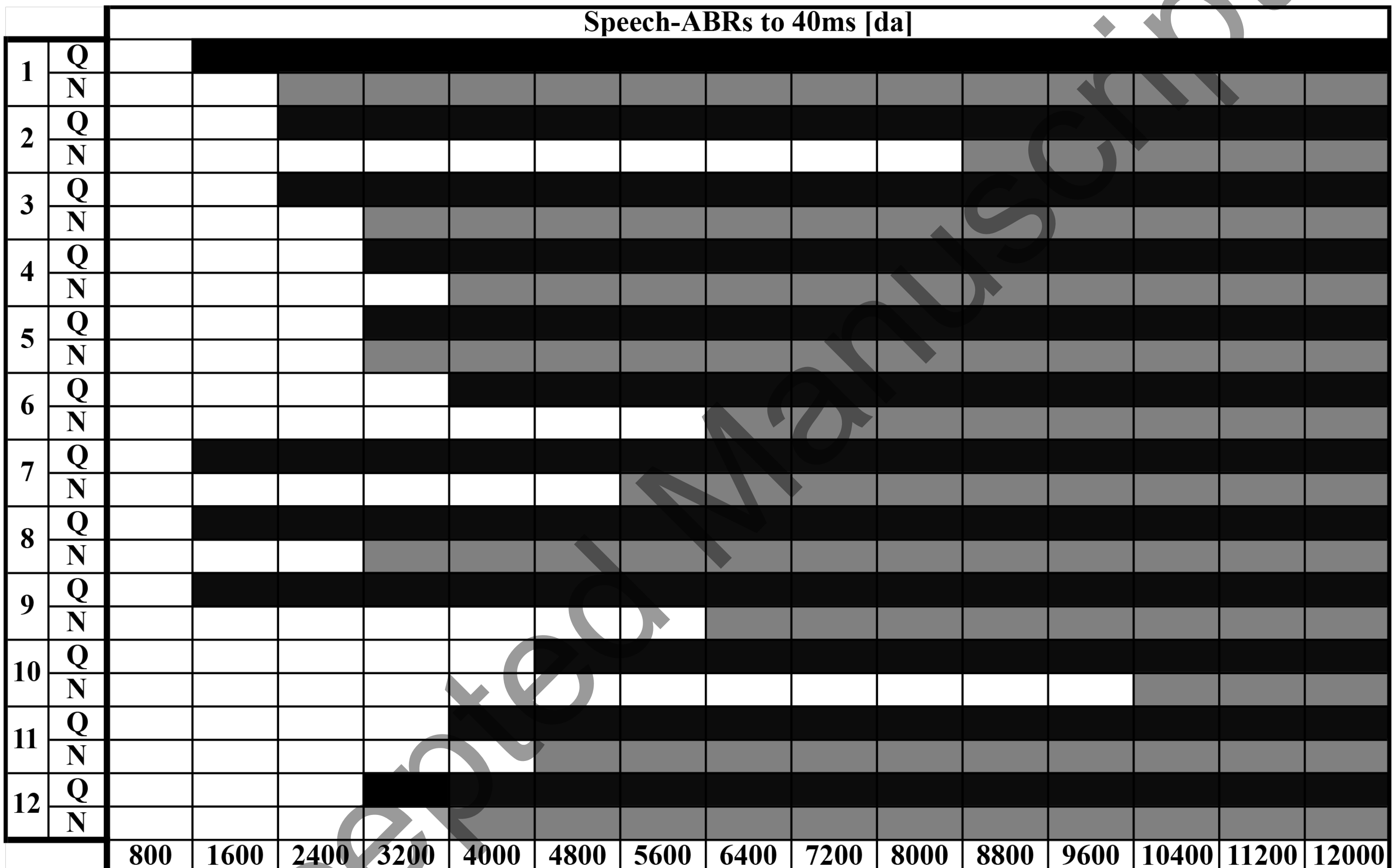


Fig. 1. Speech-ABRs to the 40ms [da] in quiet (black) and noise (grey) at 15 iterations per participant: shaded cells indicate that $F_{SP} \geq 3.1$ and all peaks that were detected at 12000 epochs were detected with 95% confidence via bootstrap. White cells indicate that not all peaks were detected with 95% confidence via bootstrap.

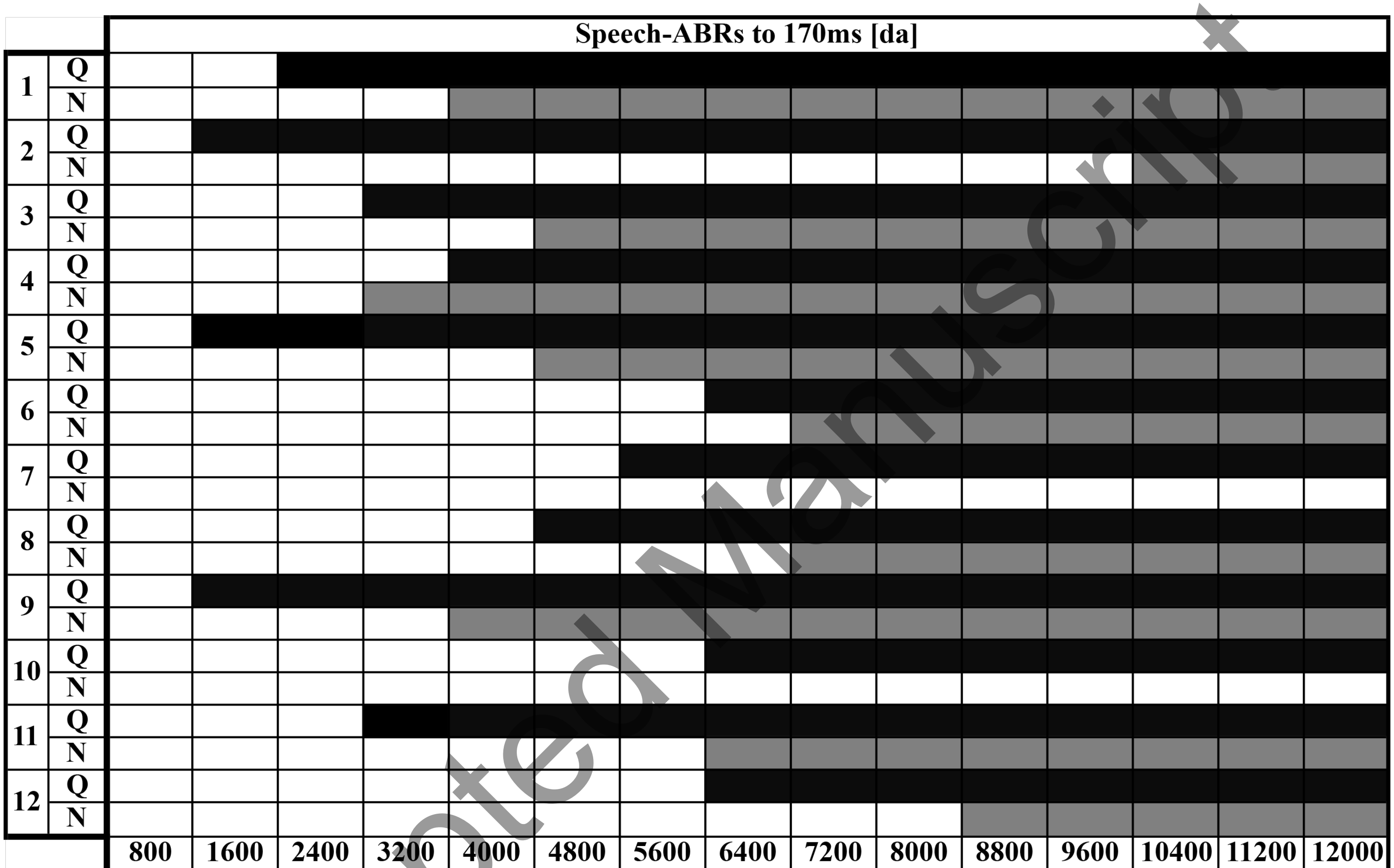


Fig. 2. Speech-ABRs to the 170ms [da] in quiet (black) and noise (grey) at 15 iterations per participant: shaded cells indicate that $F_{SP} \geq 3.1$ and all peaks that were detected at 12000 epochs were be detected with 95% confidence via bootstrap. White cells indicate that not all peaks were detected with 95% confidence via bootstrap.

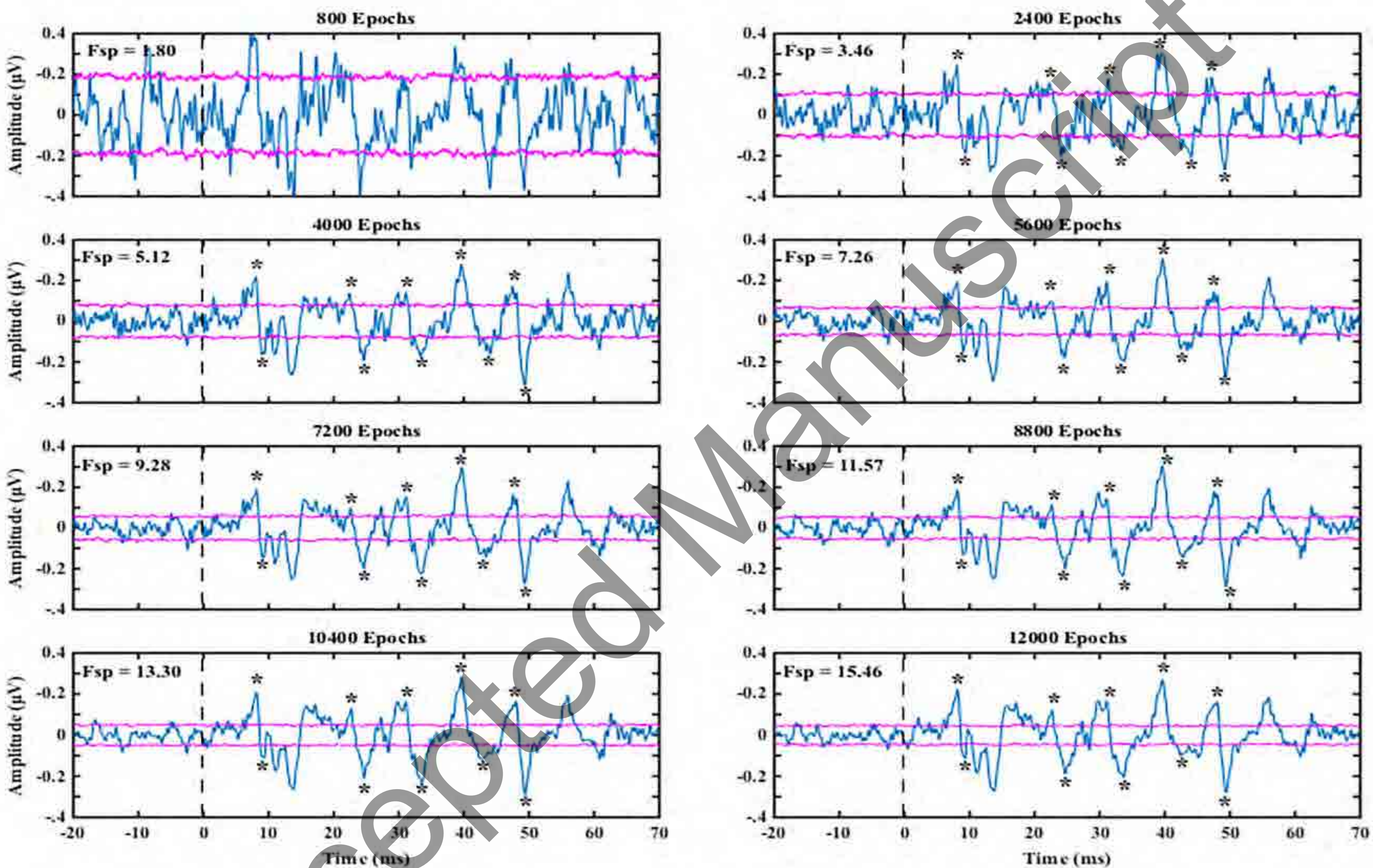


Fig. 3. Speech-ABRs with pre-stimulus baseline to the 40ms [da] in quiet at 8 iterations from a participant (5) with better responses. Peaks that were detected with 95% confidence once F_{SP} reached ≥ 3.1 are marked with a '*'.

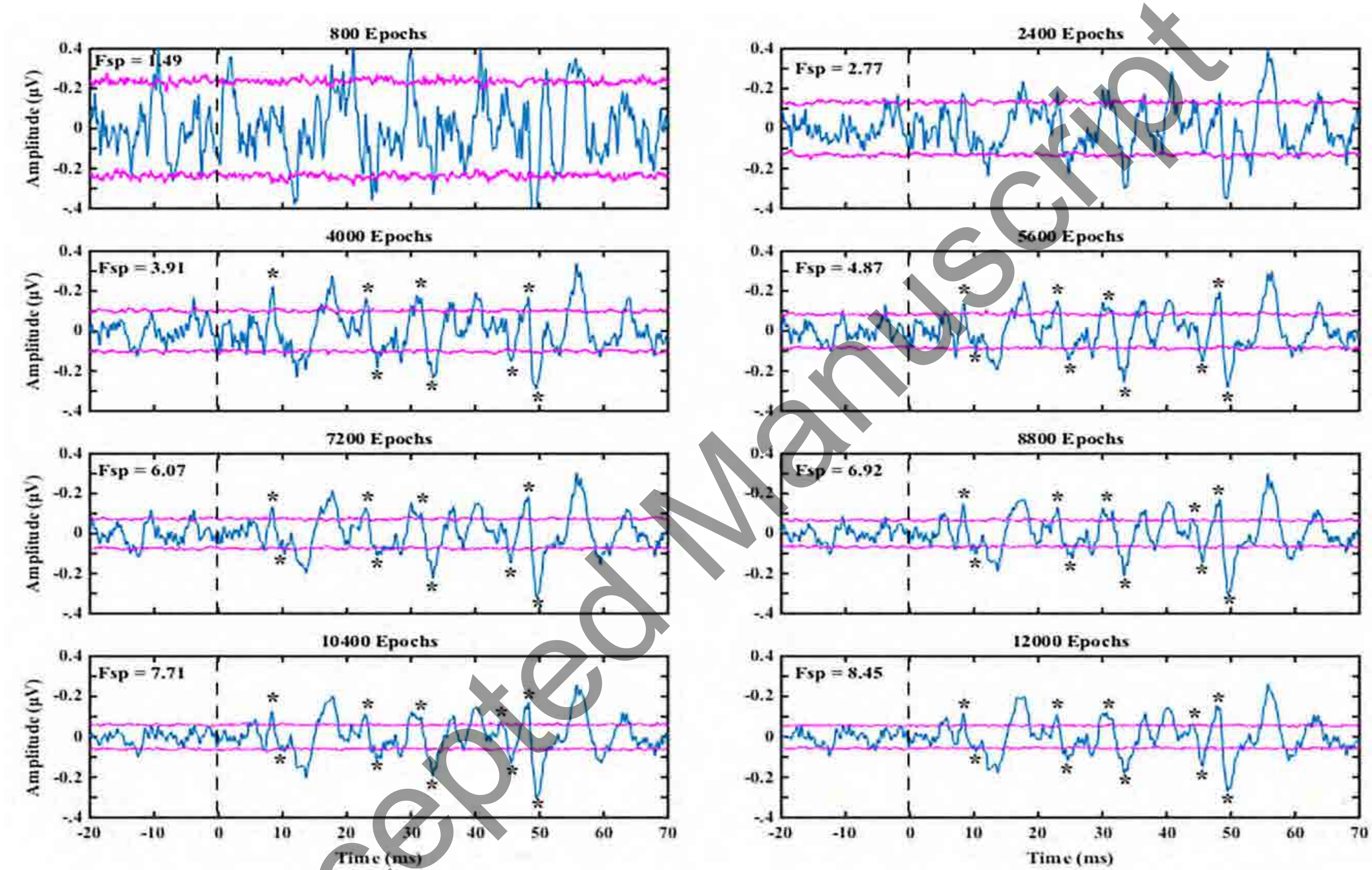


Fig. 4. Speech-ABRs with pre-stimulus baseline to the 40ms [da] in noise at 8 iterations from a participant (5) with better responses. Peaks that were detected with 95% confidence once F_{SP} reached ≥ 3.1 are marked with a '*'.

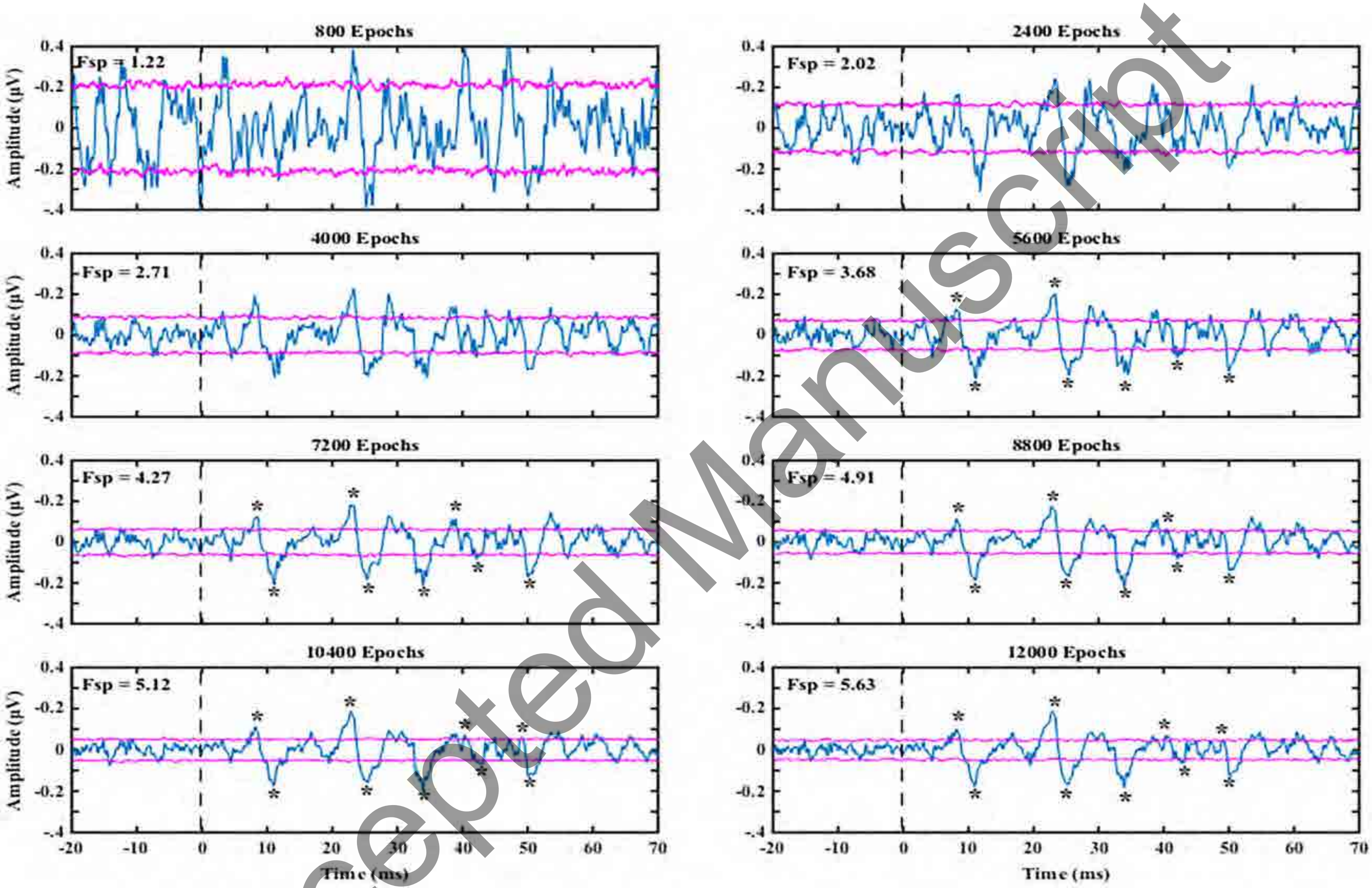


Fig. 5. Speech-ABRs with pre-stimulus baseline to the 40ms [da] in quiet at 8 iterations from a participant (10) with poorer responses. Peaks that were detected with 95% confidence once F_{SP} reached ≥ 3.1 are marked with a '*'.

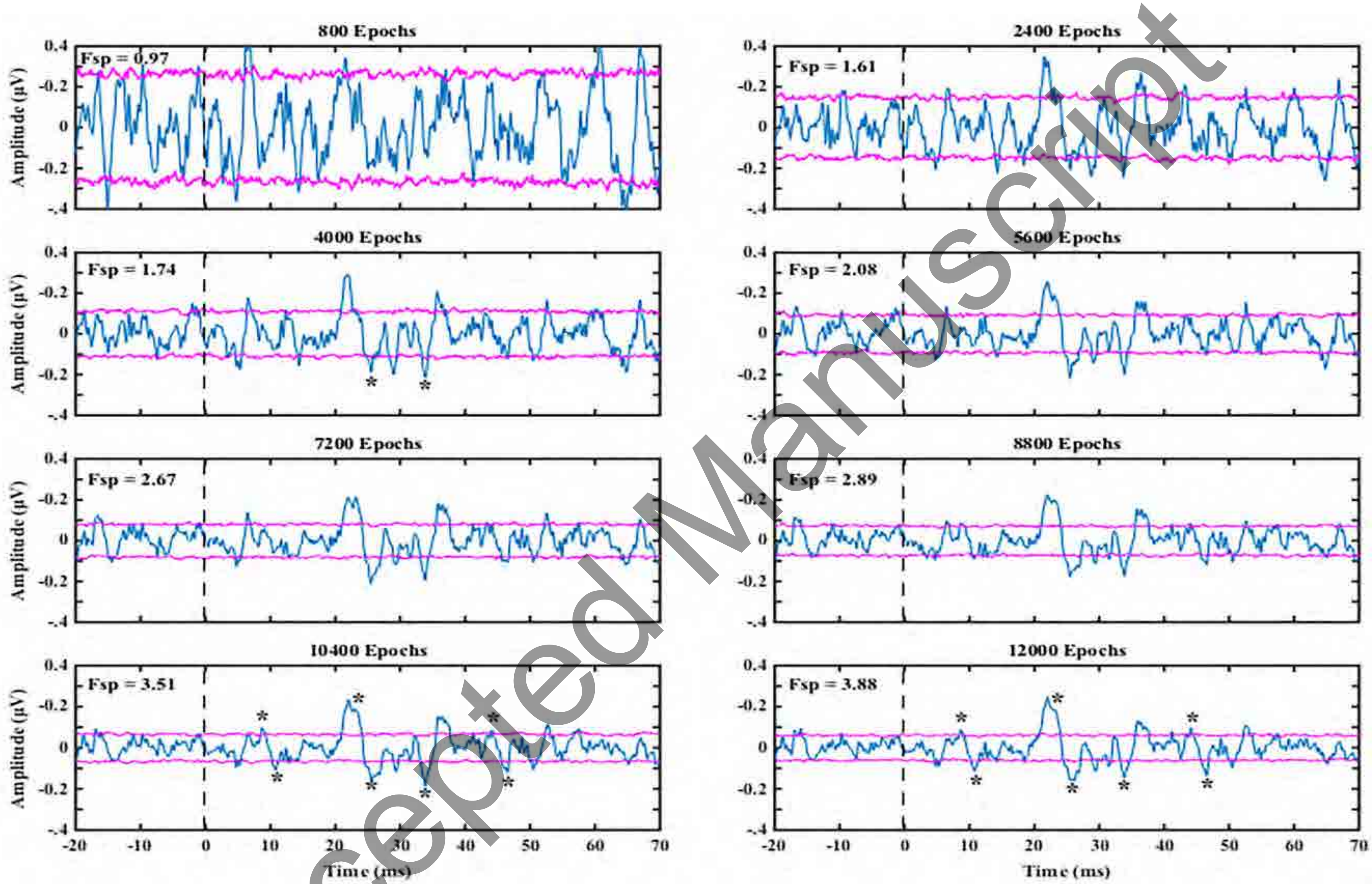


Fig. 6. Speech-ABRs with pre-stimulus baseline to the 40ms [da] in noise at 8 iterations from a participant (10) with poorer responses. Peaks that were detected with 95% confidence once F_{SP} reached ≥ 3.1 are marked with a '*'.

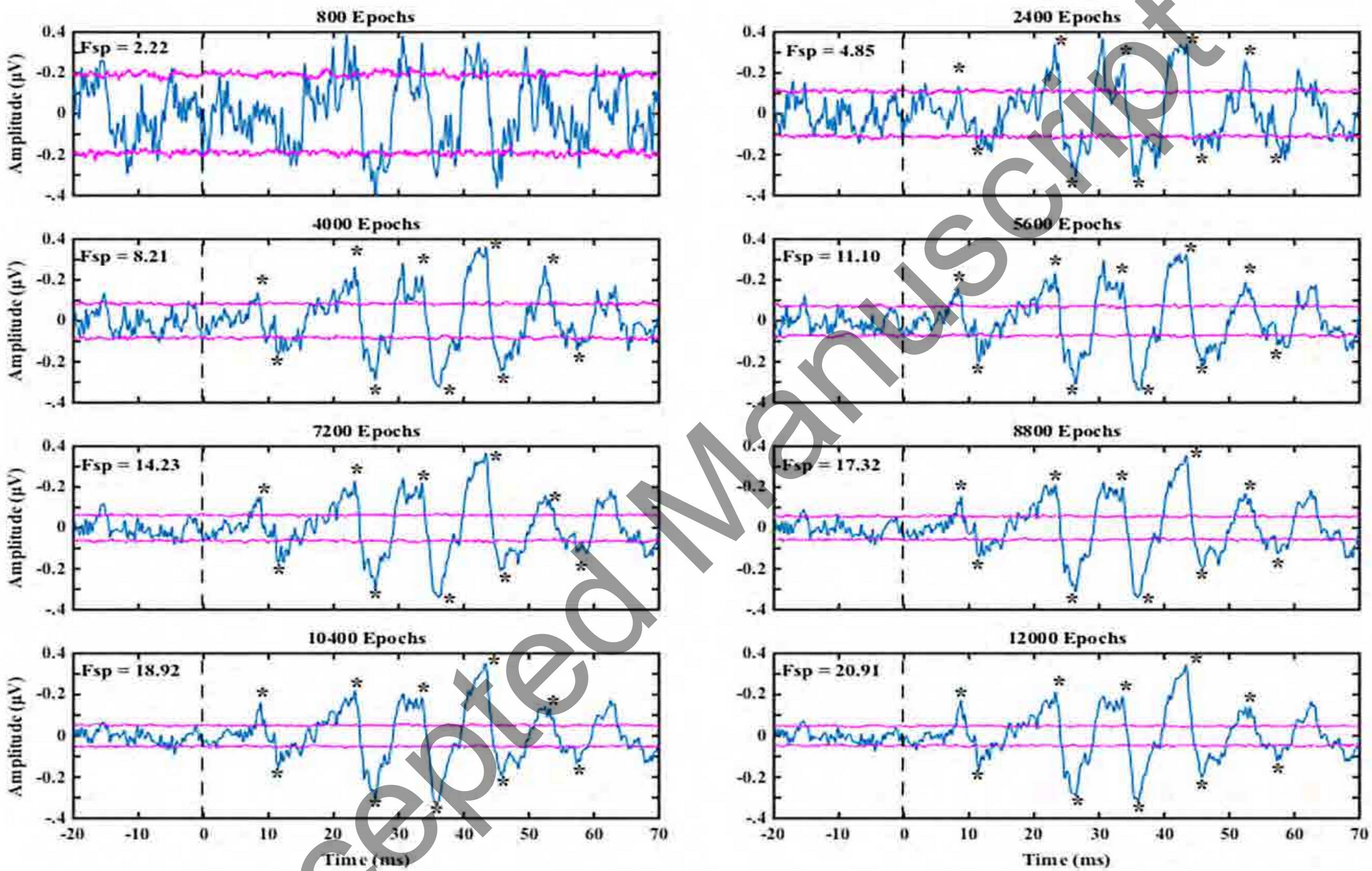


Fig. 7. Speech-ABRs with pre-stimulus baseline to the 170ms [da] in quiet at 8 iterations from a participant (5) with better responses. Peaks that were detected with 95% confidence once F_{SP} reached ≥ 3.1 are marked with a '*'.

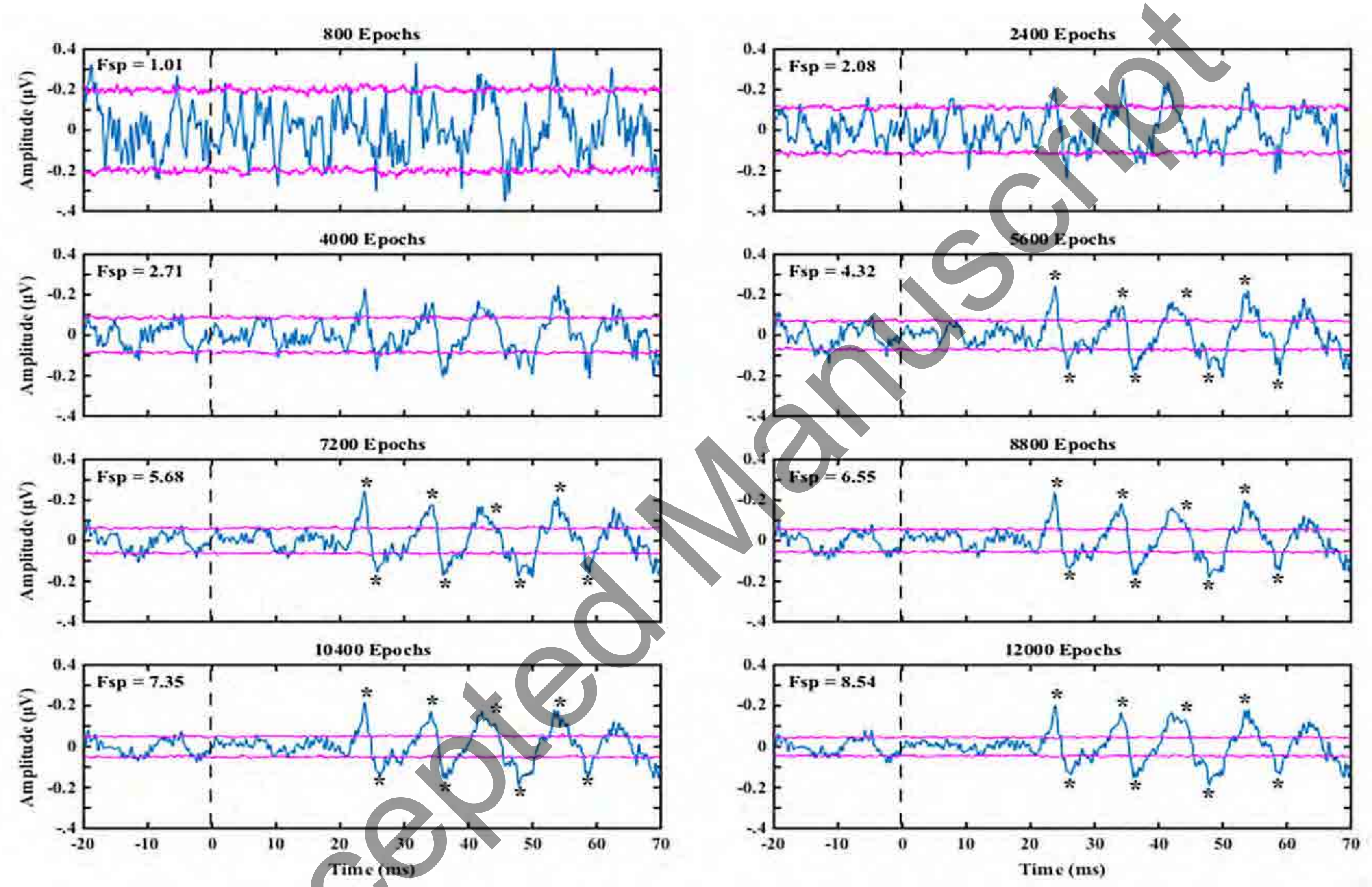


Fig. 8. Speech-ABRs with pre-stimulus baseline to the 170ms [da] in noise at 8 iterations from a participant (5) with better responses. Peaks that were detected with 95% confidence once F_{SP} reached ≥ 3.1 are marked with a '*'.

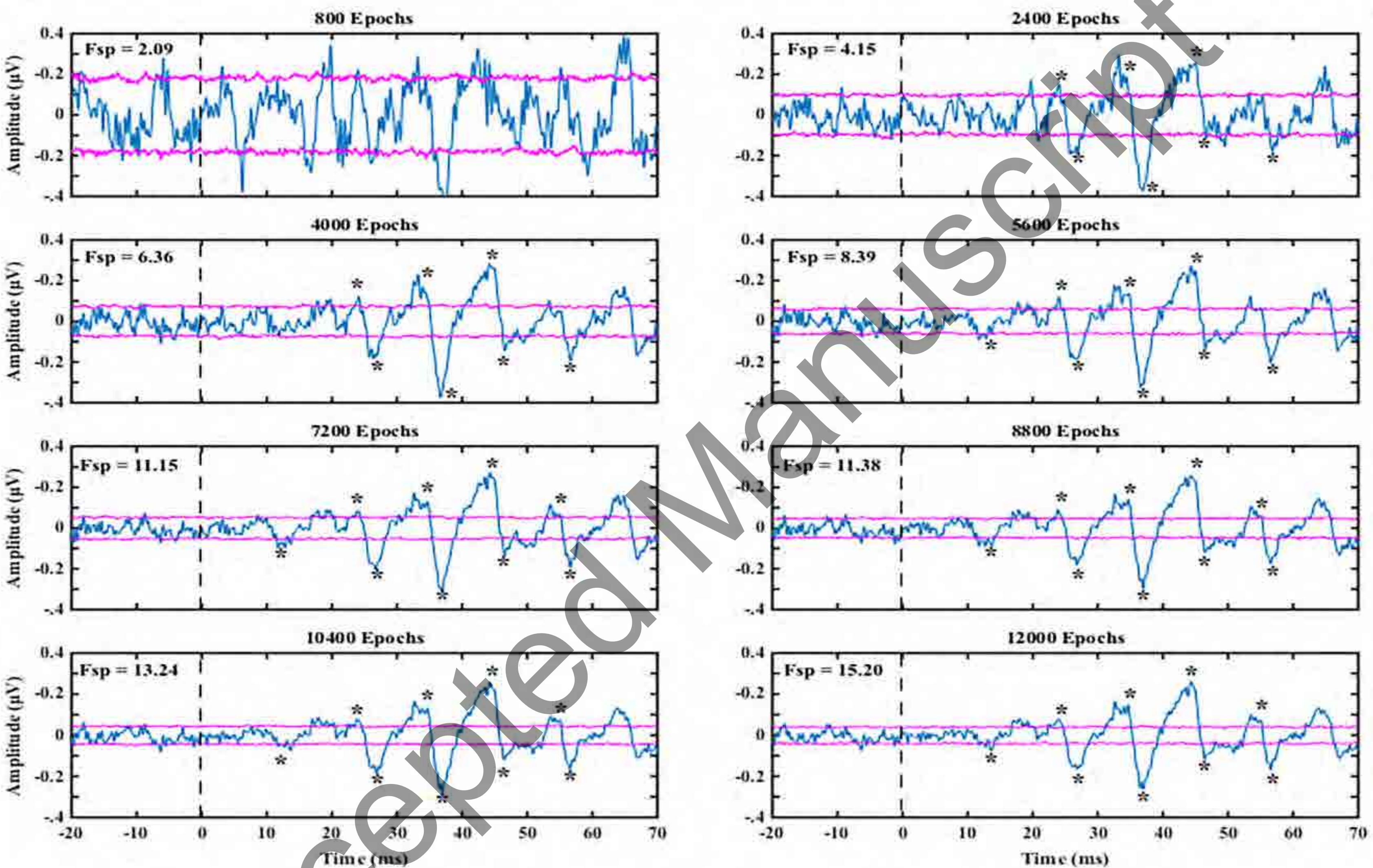


Fig. 9. Speech-ABRs with pre-stimulus baseline to the 170ms [da] in quiet at 8 iterations from a participant (10) with poorer responses. Peaks that were detected with 95% confidence once F_{SP} reached ≥ 3.1 are marked with a '*'.

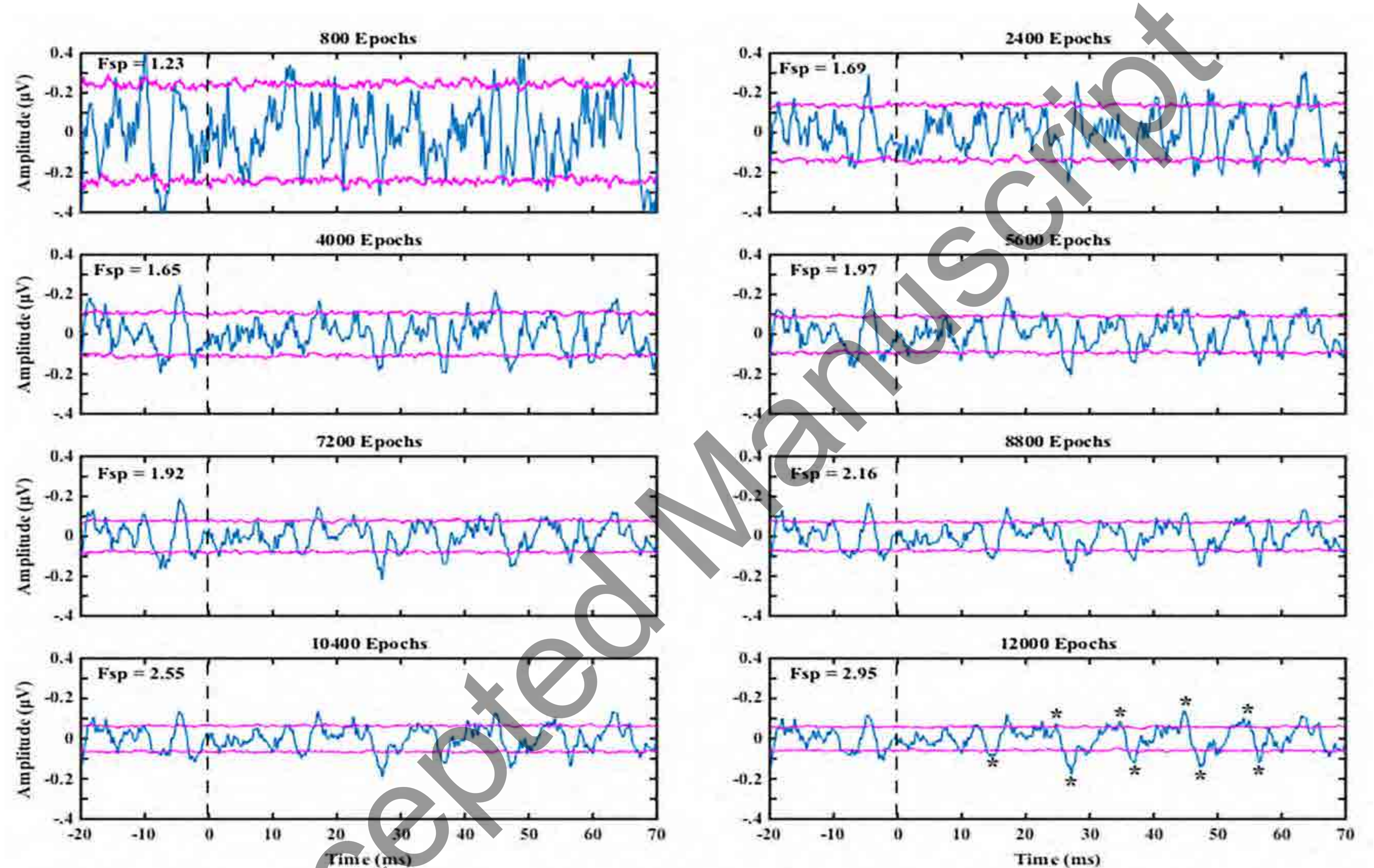


Fig. 10. Speech-ABRs with pre-stimulus baseline to the 170ms [da] in noise at 8 iterations from a participant (10) with poorer responses. F_{SP} did not reach 3.1 in this participant to the 170ms [da] in noise, therefore peaks that were detected with 95% confidence are only marked with a '*' at 12000 epochs.

Section 5: F_{SP} Values And Number of Epochs

Table 8. Number of epochs where $F_{SP} \geq 3.1$ for speech-ABRs to 40ms and 170ms [da] in quiet and in noise (per participant), F_{SP} value, difference in number of epochs between quiet and noise (diff).

Blank cells shaded in red indicate that participant speech-ABRs did not reach $F_{SP} \geq 3.1$.

	40ms [da]					170ms [da]				
	Quiet		Noise		Diff	Quiet		Noise		Diff
	Epochs	F_{SP}	Epochs	F_{SP}		Epochs	F_{SP}	Epochs	F_{SP}	
1	1600	3.37	2400	4.46	800	800	4.88	1600	3.28	800
2	2400	4.72	4800	3.83	2400	1600	3.40	5600	3.22	4000
3	2400	4.23	3200	3.34	800	3200	3.63	4000	3.23	800
4	2400	4.35	2400	3.79	0	1600	4.65	2400	3.88	800
5	2400	3.46	3200	3.58	800	1600	4.15	4800	3.50	3200
6	3200	3.50	6400	3.35	3200	1600	3.41	7200	3.44	5600
7	1600	3.50	5600	3.14	4000	2400	3.67			
8	1600	6.17	1600	3.90	0	1600	3.19	2400	3.99	800
9	1600	3.47	2400	3.13	800	1600	3.25	4000	3.21	2400
10	4800	3.16	9600	3.25	4800	2400	4.15			
11	3200	3.26	4800	3.42	1600	3200	3.32	6400	3.27	3200
12	3200	3.54	4000	3.25	800	4000	3.47	8800	3.26	4800
Mean	**2533.33**	**3.89**	**4200**	**3.54**	**1666.67**	**2133.33**	**3.76**	**4720**	**3.43**	**2181.82**
SD	**954.73**	**0.86**	**2240.13**	**0.39**	**1580.18**	**923.76**	**0.56**	**2307.86**	**0.28**	**2293.39**

Table 9. Number of epochs (at or above epochs required for $F_{SP} \geq 3.1$) where peaks were detected with 95% confidence via bootstrap for speech-ABRs to 40ms and 170ms [da] in quiet and in noise (per participant), F_{SP} values, difference in number of epochs between quiet and noise (diff).
* A larger number of epochs than required to reach $F_{sp} > 3.1$ was required to detect all peaks.
Blank cells shaded in red indicate that participant speech-ABRs did not reach $F_{SP} > 3.1$.

	40ms [da]					**170ms [da]**				
	Quiet		**Noise**		**Diff**	**Quiet**		**Noise**		**Diff**
	Epochs	**F_{SP}**	**Epochs**	**F_{SP}**		**Epochs**	**F_{SP}**	**Epochs**	**F_{SP}**	
1	1600	3.37	2400	4.46	800	2400*	12.42	4000*	6.82	1600
2	2400	4.72	8800*	6.16	6400	1600	3.40	10400*	4.75	8800
3	2400	4.23	3200	3.34	800	3200	3.63	4800*	3.71	1600
4	3200*	5.46	4000*	5.07	800	4000*	10.34	3200*	4.84	-800
5	2400	4.28	3200	3.58	800	1600	4.15	4800	3.5	3200
6	4000*	4.12	6400	3.35	2400	6400*	11.5	7200	3.44	800
7	1600	3.50	5600	3.14	4000	5600*	6.36			
8	1600	6.17	3200*	5.56	1600	4800*	10.51	7200*	8.86	2400
9	1600	3.47	6400*	5.98	4800	1600	3.25	4000	3.21	2400
10	4800	3.16	10400*	3.51	5600	6400*	9.73			
11	4000*	3.99	4800	3.42	800	3200	3.32	6400	3.27	3200
12	3200	3.54	4000	3.25	800	6400*	4.66	8800	3.26	2400
Mean	**2733.33**	**4.17**	**5200.00**	**4.24**	**2466.67**	**3933.33**	**6.94**	**7066.67**	**4.30**	**3133.33**
SD	**1103.16**	**0.91**	**2448.00**	**1.15**	**2142.78**	**1943.44**	**3.65**	**3123.32**	**1.82**	**2630.36**

Table 10. F_{sp} values for speech-ABRs to the three [da] durations in quiet and in noise (per participant) at 12000 epochs and 'no sound' F_{sp} values (per participant).

F_{sp} values in red are those below 3.1

	40ms [da]		50ms [da]		170ms [da]		No Sound
	Quiet	**Noise**	**Quiet**	**Noise**	**Quiet**	**Noise**	
1	23.69	19.00	23.06	18.63	22.32	17.83	1.05
2	15.84	9.15	22.41	4.47	10.74	5.09	0.66
3	16.00	9.30	8.34	13.41	8.22	11.42	0.82
4	20.87	11.96	17.88	13.70	5.51	13.83	1.27
5	15.46	8.45	7.86	5.58	10.94	8.54	0.70
6	10.99	5.01	13.93	5.22	7.29	5.15	1.02
7	20.04	6.95	9.82	3.46	4.64	**2.96**	0.76
8	36.69	19.99	21.96	11.38	11.64	14.45	1.05
9	25.71	10.04	16.67	9.62	8.48	8.81	1.46
10	5.63	3.88	6.76	4.29	6.37	**2.95**	0.72
11	11.12	5.86	3.68	**2.76**	5.45	5.34	0.87
12	12.84	8.14	9.36	4.24	3.60	4.28	1.09
Mean	**17.91**	**9.81**	**13.48**	**8.06**	**8.77**	**8.39**	**0.95**
SD	**8.23**	**5.05**	**6.76**	**5.14**	**4.99**	**4.97**	**0.25**

Table 11. F_{sp} values for speech-ABRs (per participant) to the 50ms [ba] and [ga] in quiet and in noise and to the 170ms [ba] and [ga] in quiet at 12000 epochs.

F_{sp} values in red are those below 3.1

	50ms [ba]		170ms [ba]	50ms [ga]		170ms [ga]
	Quiet	**Noise**	**Quiet**	**Quiet**	**Noise**	**Quiet**
1	50.91	24.51	48.21	18.15	8.85	29.36
2	8.11	**3.06**	29.92	5.92	**1.79**	31.24
3	11.05	3.27	16.12	4.17	**2.83**	10.82
4	8.81	7.55	28.48	11.68	10.98	25.38
5	11.55	4.61	21.60	8.88	4.95	25.37
6	13.47	4.31	17.36	11.90	5.09	13.90
7	9.89	7.03	12.22	10.60	3.98	15.74
8	29.63	13.60	24.38	20.99	11.21	19.36
9	17.48	12.44	37.67	19.73	8.78	41.69
10	7.19	**2.37**	8.52	6.77	3.56	12.66
11	6.64	4.12	6.92	5.38	3.44	7.38
12	6.42	4.36	15.32	11.25	2.86	12.76
Mean	**15.10**	**7.60**	**22.23**	**11.29**	**5.69**	**20.47**
SD	**12.99**	**6.43**	**12.24**	**5.67**	**3.34**	**10.20**