A Time-Saving Alternative to "Peak-Picking" Algorithms: A Gaussian Mixture Model Feature Extraction Technique for the Neurodiagnostic Auditory Brainstem Response

Aryn M. Kamerer

Objectives: The accurate and efficient analysis of neurodiagnostic auditory brainstem responses (ABR) plays a critical role in assessing auditory pathway function in human and animal research and in clinical diagnosis. Traditional analysis of the neurodiagnostic ABR analysis involves visual inspection of the waveform and manually marking peaks and troughs. Visual inspection is a tedious and time-consuming task, especially in research where there may be hundreds or thousands of waveforms to analyze. "Peak-picking" algorithms have made this task faster; however, they are prone to the same errors as visual inspection. A Gaussian mixture model-based feature extraction technique (GMM-FET) is a descriptive model of ABR morphology and an alternative to peak-picking algorithms. The GMM-FET is capable of modeling multiple waves and accounting for wave interactions, compared with other template-matching approaches that fit single waves.

Design: The present study is a secondary analysis applying the GMM-FET to 321 ABRs from adult humans from 2 datasets using different stimuli and recording parameters. Goodness-of-fit of the GMM-FET to waves I and V and surrounding waves, that is, the summating potential and waves IV and VI, was assessed, and latency and amplitude estimations by the GMM-FET were compared with estimations from visual inspection.

Results: The GMM-FET had a similar success rate to visual inspection in extracting peak latency and amplitude, and there was low RMS error and high intraclass correlation between the model and response waveform. Mean peak latency differences between the GMM-FET and visual inspection were small, suggesting the two methods chose the same peak in the majority of waveforms. The GMM-FET estimated wave I amplitudes within 0.12 μV of visual inspection, but estimated larger wave V amplitudes than visual inspection.

Conclusions: The results suggest the GMM-FET is an appropriate method for extracting peak latencies and amplitudes for neurodiagnostic analysis of ABR waves I and V.

Key words: Auditory brainstem response, Feature extraction, Gaussian mixture model, Peak-picking.

Abbreviations: ABR = auditory brainstem response; AP = action potential; GMM-FET = Gaussian mixture model-based feature extraction tool; HL = hearing loss; ICC = intraclass correlation; IQR = interquartile range; LoA = limits of agreement; Mi = mild HL; NH = normal hearing; NRMSE = normalized root-mean-square-amplitude; PST = poststimulus time; SP = summating potential; WI-sub. = within subject.

(Ear & Hearing 2024;45;1115–1124)

Department of Communicative Disorders and Deaf Education, Utah State University, Logan, Utah, USA.

INTRODUCTION

The auditory brainstem response (ABR) is a series of sound-evoked potentials characterized by a series of reliable peaks generated by compound action potentials (AP) of ascending auditory ganglia and nuclei. The cellular generators of these peaks have been established in animal models (Melcher & Kiang 1996; Melcher et al. 1996a, b), but are less distinguished in humans. Regardless of species, it is generally agreed that wave I (or 1 in animals) is generated by the auditory nerve and subsequent waves are generated by brainstem nuclei. Because of the reliability of these peaks, the ABR is used as an objective measure of human hearing in both research and clinical settings and is a ubiquitous measure of auditory function in animal research. ABRs can be used to either estimate hearing sensitivity (i.e., threshold ABR) or to identify possible pathophysiology (i.e., site-of-lesion or neurodiagnostic ABR) (Coats & Martin 1977; Galambos & Hecox 1978; Jerger & Mauldin 1978; Gorga et al. 1985). While analysis of the threshold ABR determines whether a response is present at a given stimulus level, analysis of the neurodiagnostic ABR is more involved and requires measuring waveform features and interpreting whether those measurements fall within a normative range or are indicative of pathology.

Analysis of the neurodiagnostic ABR has traditionally relied on visual inspection of the waveform, and manually clicking or placing markers on features such as peaks, troughs, and baselines, from which metrics such as peak latency and amplitude are calculated. The task can be quite tedious and time-consuming, especially for research studies that may collect thousands of response waveforms. In addition to time, reliance on human judgment to analyze the waveform also injects subjectivity and human error into an objective measure of auditory physiology (Vidler & Parker 2004). Choosing peaks by eye can be especially difficult when noise is introduced either internally (e.g., muscle contraction) or externally (e.g., electrical artifact); when morphology is altered by pathology; or when waves overlap in time. Wave overlap is a common morphological occurrence that can obscure the visual landmarks required for neurodiagnostic measurement. For example, wave V is an important wave for human neurodiagnostic purposes, but it often overlaps with waves IV and VI (Fig. 1A). The most common way to estimate peak amplitude is from the peak maximum to the following minimum, as in Figure 1A "I" to "I_{trough}." This method is difficult to implement on wave V in Figure 1A example. The onset of wave VI often overlaps with the offset of wave V, eliminating the wave V minimum or "trough." It is common in visual inspection protocols to circumvent this problem by placing the wave V trough marker at the shoulder where wave V meets the rising

<zdoi; 10.1097/AUD.0000000000001498>

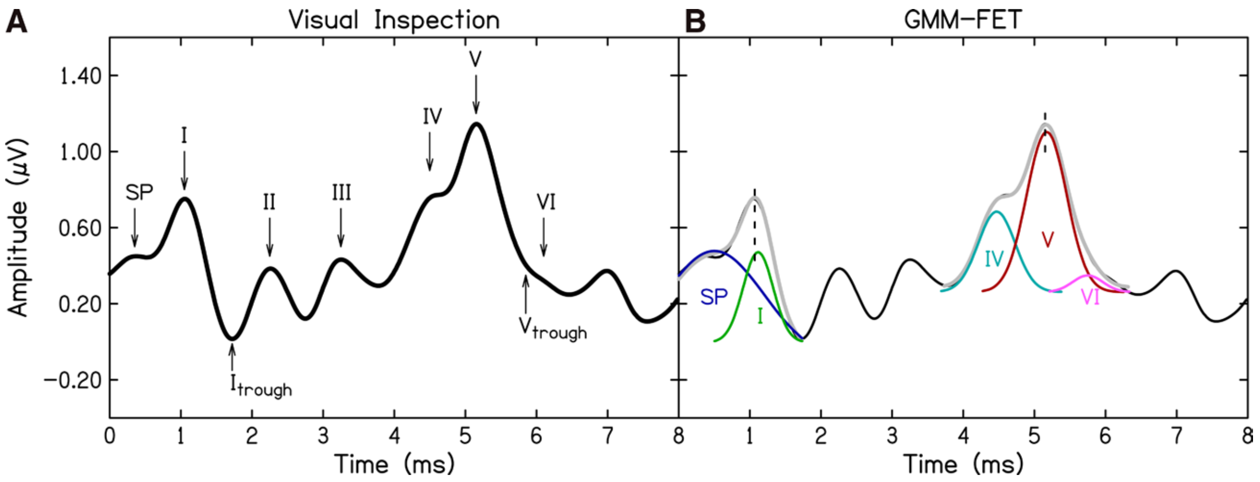

Fig. 1. Auditory brainstem response waveform from an adult human (solid black lines). A, The traditional method of analysis by visual inspection of waveform features and manual marking of peaks and troughs (indicated by arrows). B, The same ABR as (A) but analyzed using the GMM-FET, that fits a mixture of individual Gaussian functions (colors) that summed (gray) model any portion of the waveform. Dashed black lines mark peak latencies for waves I and V estimated by visual inspection to highlight differences in latency estimations between GMM-FET and visual. ABR indicates auditory brainstem response; GMM-FET, Gaussian mixture model feature extraction technique; SP, summating potential.

wave VI (Fig. 1A, "V_{trough}"); however, this compromise results in an underestimation of wave V amplitude.

Computational analysis of the waveform may solve some problems of visual inspection. A great number of response detection algorithms have been created for threshold ABRs that detect the response (Elberling 1979; Arnold 1985; Sininger 1993; Delgado & Ozdamar 1994; Don & Elberling 1994; Bradley & Wilson 2004; Irino & Patterson 2009; Suthakar & Liberman 2019), with some embedded into infant hearing screening equipment. On the other hand, there are no widely accepted computational analyses for neurodiagnostic ABR analysis. Many research laboratories have written their own "peak-picking" algorithms, which effectively locate local maxima and following minima within a specified latency range. Although these algorithms save time by providing initial estimates, there are two drawbacks. The first is that these algorithms must be tailored to the dataset, as latency ranges can vary based on test parameters such as electrode montage and stimulus rate, and auditory disorders such as loss of hearing sensitivity. The second is that these algorithms mimic the process of visual inspection and are therefore prone to the same weaknesses, such as difficulty determining the correct peaks in noisy waveforms and finding troughs in overlapping waves. Statistical alternatives to peak-picking algorithms have been proposed that do not suffer from the same drawbacks. Examples of alternatives include dynamic time-warping (Picton et al. 1988; Krumbholz et al. 2020) and template-matching (Elberling 1979; Vannier et al. 2002; Chertoff 2004; Valderrama et al. 2014). These approaches model individual waves and can extract features such as peak latency and amplitude, as well as additional features that are difficult to detect visually, such as wave width and area. Kamerer et al. (2020) recently developed a similar analytical technique to extract features from the ABR waveform, but one that fits multiple waves and accounts for interactions between waves. The analytical technique implements a Gaussian mixture model (GMM) fitting individual waves in series to recreate any portion of the waveform. The GMM works as a descriptive model of waveform morphology, rather than physiology. ABR waves reflect the poststimulus time (PST) histogram of AP from a large population of auditory ganglia or nuclei, which are biphasic in shape rather than Gaussian (Goldstein & Kiang 1958). But recorded from the scalp, the ABR is a far-field potential, and the biphasic shape of the compound neural potentials are lost due to filtering by the head tissues and overlap with the oncoming wave from the next neural generator, leaving behind a coincidental series of Gaussian-shaped waves. From the GMM, features such as peak latency, peak amplitude, onset (or first-spike) latency, and wave width, can be extracted; hence, we have named it the GMM feature extraction technique or GMM-feature extraction technique (FET). The GMM-FET was developed and validated on the summating potential (SP) and wave I (or "wave I complex") of adult human ABRs (Kamerer et al. 2020). Their study found an excellent fit of the GMM-FET to the wave I complex and high agreement between wave I amplitudes estimated by the GMM-FET and amplitudes estimated by visual inspection.

In the present study, we expand the GMM-FET to include waves IV, V, and VI of the ABR and assess its ability and ease to generalize to new data. These waves were chosen because of the importance of waves I and V for human neurodiagnostic purposes and because wave V often overlaps with waves IV and VI, visual inspection is prone to inaccurate measures of peak amplitude. In the present study, we assess goodness-of-fit of the GMM-FET to the SP and waves I, IV, V, and VI of two large datasets of ABR waveform responses, and compare peak latency and amplitude estimations of waves I and V to those estimated by visual inspection.

MATERIALS AND METHODS

Participants

Dataset 1 • ABR recordings were collected from 221 adults ages 20 to 86 (mean age = 45; 119 female). Portions of this dataset were previously published (Kamerer et al., 2019a, b) or are part of a forthcoming publication. All participants had pure-tone air conduction thresholds (HD 200 over-the-ear headphones; Sennheiser, Wedemark, Germany) of ≤65 dB HL at all octave frequencies (0.25 to 8 kHz) and 3 interoctave frequencies

(0.75, 3, and 6 kHz). Furthermore, participants had no evidence of middle ear dysfunction as inclusion to the study required middle ear pressure within the range of −100 to +50 daPa and static compliance between 0.3 and 2.5 cm^3 measured via 226-Hz tympanometry (Otoflex 100, Madsen) and no air-bone gap >10 dB at any octave frequency from 500 to 4000 Hz. In the original study, the 99 participants had audiometric thresholds at all octave frequencies (0.25 to 8 kHz) ≤15 dB HL and were considered to have normal hearing. The other 122 participants with thresholds >15 dB HL were considered to have hearing loss. ABR recordings were made monaurally in the better ear. If there was no apparent better ear, the ear was chosen at random though efforts were made to have equal numbers of left and right ears.

Dataset 2 • ABR recordings were collected from 100 adults ages 19 to 55 (mean age = 36; 71 female). These data were collected for a forthcoming publication. Participants in this dataset had pure-tone air conduction thresholds (HD 200 over-the-ear headphones; Sennheiser, Wedemark, Germany) of ≤40 dB HL at all octave frequencies (0.25 to 8 kHz) and 3 inter-octave frequencies (0.75, 3, and 6 kHz). Participants also no evidence of middle ear dysfunction as inclusion to the study required middle ear pressure within the range of −100 to +50 daPa and static compliance between 0.3 and 2.5 cm^3 measured via 226-Hz tympanometry (Otoflex 100, Madsen) and no air-bone gap >10 dB at any octave frequency from 500 to 4000 Hz. Part of the inclusion criteria for this study was air conduction symmetry between ears (<10 dB difference in 4-frequency PTA); therefore, the ear for ABR recordings was chosen at random though efforts were made to have equal numbers of left and right ears.

Auditory Brainstem Response

Both datasets used custom software (Cochlear Response [CResp] version 1.0; Omaha, NE) on a computer equipped with a 24-bit soundcard (Babyface Pro; RME, Germany). Calibration of the stimulus levels was done using a sound level meter (System 824 and SoundTrack LxT1; Larson Davis, Provo, UT) with the ER3 connected to the sound level meter via a 2cc coupler (G.R.A.S. 60126, Denmark).

Dataset 1 • Tone burst-elicited ABR waveforms were recorded at 4 kHz using surface electrodes placed at the forehead (Fpz, ground), vertex (Cz, noninverting active), and an inverting reference electrode placed in the ear canal (ER3-26A gold foil tiptrodes). The 4 kHz pure tone was gated via Blackman window with duration of 1 msec. Stimuli were presented in alternating polarity monaurally at a rate of 11/sec to an ER-3A insert earphone (Etymotic Research, Elk Grove, IL) connected to the soundcard. The stimulus sound-pressure level (SPL) was 110 dB peak-equivalent (pe)SPL. Electrode impedances were ≤5 kΩ in all cases. The electroencephalography (EEG) signal was amplified (gain = 100,000), filtered 0.1 to 1.5 kHz and for line interference using a 60 Hz notch filter (Opti-Amp 8001; Intelligent Hearing Systems, Miami, FL) and directed to the computer via the soundcard for averaging. Responses were separated by even and odd recordings and stored in two buffers which were averaged for the final waveform (total averages = 1500 artifact-free responses). Artifact rejection was based on the peak absolute differences between the buffers and was set at a maximum of ±20 μV.

Dataset 2 • Click-elicited ABR waveforms were recorded using surface electrodes placed at the forehead (Fpz, ground), vertex (Cz, noninverting active), and an inverting reference electrode placed in the ear canal (ER3-26A gold foil tiptrodes). The click-stimulus was 100 μsec, presented in alternating polarity monaurally at a rate of 11.3/sec to an ER-3A insert earphone (Etymotic Research, Elk Grove, IL) connected to the soundcard at a SPL of 100 dB peSPL. Electrode impedances were ≤5 kΩ in all cases. The EEG signal was amplified (gain = 100,000), filtered 0.01 to 3000 Hz and for line interference using a 60 Hz notch filter (Opti-Amp 8001; Intelligent Hearing Systems, Miami, FL) and directed to the computer via the soundcard for averaging. Responses were separated by even and odd recordings and stored in 2 buffers which were averaged for the final waveform (total averages = 1000 artifact-free responses). Artifact rejection was based on the peak absolute differences between the buffers and was set at a maximum of ±20 μV.

Visual Inspection

Data from visual inspection of peaks and troughs were only available for dataset 1. The software allowed for a resolution of 0.02 μV for amplitude and 0.02 msec for latency. Wave amplitude was calculated as the difference between the positive peak and the following trough. The processing delay of the soundcard was considered when analyzing the data for latency. Two examiners independently identified peaks and troughs of ABR waves I and V. Peak or trough latency differences >0.02 msec, which occurred in 37 of the total 221 waveforms (~17%), were resolved by a third examiner. Of the 37 waveforms with disagreements, there were 9 disagreements on wave I peak latency, 11 disagreements on wave I trough latency, 12 disagreements on wave V peak latency, and 33 disagreements on wave V trough latency.

Gaussian Mixture Model Feature Extraction Technique

Waveforms were preprocessed for analysis using MATLAB. Waveforms were bandpass filtered using a zero-phase infinite impulse response filter with cutoff frequencies of 60 and 1500 Hz. Each waveform was DC-offset adjusted by calculating the mean amplitude in a 500 μsec window around time zero and subtracting that value from the entire waveform. The ABR was windowed from 0 to 2 and 3.5 to 7 msec PST for the wave I and wave V complexes, respectively. This windowing was performed to exclude the stimulus artifact and other waves. The ABR waveform amplitude as a function of time was modeled by a Gaussian mixture:

$$\mathrm{ABR}\,(t) = \mathrm{DC} + A_1 e^{\frac{(t-L_1)^2}{2W_1^2}} + \ldots + A_n e^{\frac{(t-L_n)^2}{2W_n^2}},$$

where each Gaussian models an individual wave. DC is the offset adjustment around time zero, t is PST (msec), A is peak amplitude (V), L is peak latency (msec), and W is semiwidth (msec). A least-squares nonlinear regression analysis estimated 3 parameters per wave for 15 parameters total per waveform. Initial values and boundary conditions were set for all parameters to speed estimation and convergence of the algorithm and ensure accurate wave identification. The initial value for the peak latency (L) parameter was set manually for each wave based on looking at several waveforms in each dataset and assigning a value that captured the general location of peaks. The bound constraints for the peak latency (L) parameter were originally set to ±0.5 around the initial value. After running the model on the dataset, bound constraints were tightened if the

GMM-FET erroneously chose the incorrect peak to fit. Search bounds were set to $0 \leq A \leq 5$ μV and $0.2 \leq W \leq 0.7$ msec for all waves. Initial values and search bounds for latency (L) are shown in Table 1.

Statistical Analysis

The performance of the model was assessed in two ways: (1) ability of the Gaussian mixture to model morphology, and (2) agreement between the Gaussian mixture and visual inspection on peak latency and amplitude. The ability of the Gaussian mixture to model waveform morphology was measured by model fit to the ABR in the time-domain, using intraclass correlation (ICC) and root-mean-square error of the residuals normalized to the maximum amplitude of the waveform (NRMSE). While both measure goodness-of-fit, ICC is a measure of agreement while NRMSE is a measure of the differences between the model and observed data. Together they provide a complete picture of the goodness-of-fit. The success rate of the GMM-FET was determined as the percentage of waveforms with good or excellent agreement between the GMM-FET and waveform, defined as ICC ≥ 0.75 (Koo & Li 2016). GMM-FET models with moderate or poor agreement with the waveform (ICC < 0.75) were considered unsuccessful. These numbers are reported. The GMM-FET analysis was completed in MATLAB along with the ICC calculations (Salarian 2016). The agreement between the Gaussian mixture and visual inspection was performed on dataset 1 and quantified via Bland-Altman analysis (Bland & Altman 1999). While correlational analyses study the relationship between two methods, they do not quantify the differences. Bland-Altman is way to compare two methods that analyzes the mean difference between two methods and quantifies method bias. The Bland-Altman analyses was completed in R.

RESULTS

The GMM-FET was asked to fit a GMM to the wave I complex that includes the SP and wave I, and the wave V complex that includes waves IV-VI of the ABRs of 321 adults. Dataset 1 included 221 responses from adults with both normal hearing and hearing loss and dataset 2 included 100 responses from adults with normal hearing and mild hearing loss. The model was assessed via goodness-of-fit to the response waveforms for each complex in datasets 1 and 2. In addition, for dataset 1, the GMM-FET was compared with visual inspection of wave amplitude via Bland-Altman analysis. A sample of 16 waveforms (gray) with the GMM-FET model (red) are presented in Figure 2.

Goodness-of-Fit

Of 642 total waveforms (321 wave I and 321 wave V), the success rate, or percentage of waveforms fit with ICC ≥ 0.75, was 99%. Success was dependent on wave complex, with 98% (314) wave I complexes and 100% (321) wave V complexes successfully modeled (Table 2). Success was also dependent on dataset and hearing status. The success rate for the wave I complex was 98% for normal-hearing participants in dataset 1, 96% for participants with hearing loss in dataset 1, and 100% for normal hearing to mild hearing loss in dataset 2. In total, there were 7 wave complexes (of 642) that were poorly fit, all of which were wave I, and 5 of which came from participants with hearing loss. The ICC and NRMSE for each wave complex are shown in Figure 3. While there was a significant negative correlation between ICC and NRMSE ($R = -0.52$, $p < 0.001$), there were instances when ICC was high but NRMSE was relatively high (i.e., poor), or vice versa, for a given wave complex (Fig. 2, right panel). High ICC with high NRMSE generally meant the wave shape was captured by the GMM-FET but was slightly shifted in time or amplitude, while low ICC and low NRMSE typically meant the waveform was noisier and the extra bumps in the waveform were not captured by the model, or not all waves were captured well (Fig. 2, right panel). Median NRMSE across datasets was 3% for wave I (Fig. 3, black circle) and 0.5% for wave V (Fig. 3, black triangle). Median ICC was 0.996 for wave I and 0.999 for wave V. While there is excellent goodness-of-fit overall, there was better fit to the wave V complexes than wave I as seen in the distributions of ICC and NRMSE in Figure 3.

Agreement With Visual Inspection

Peak latencies and amplitudes for waves I and V in dataset 1 had been visually inspected by two expert audiologists (with a third acting as arbitrator) as part of the study for which the ABRs were collected. These latencies and amplitudes were compared with those estimated by the GMM-FET via Bland-Altman analysis of agreement (Fig. 4). The seven waveforms that were not modeled successfully in this dataset were removed before agreement analysis. There were two wave Is and one wave V where the visual inspectors fail to determine an amplitude but the GMM-FET was able to model. These three waves were also removed from the dataset before agreement analysis. Agreement was affected by the hearing status of the participants; therefore, Bland-Altman analyses were performed separately on participants with normal hearing, defined as audiometric thresholds ≤15 dB HL (n = 97 wave I and 99 wave V) and those with audiometric hearing loss (thresholds >15 dB HL; n = 117 wave I and 122 wave V).

As expected, mean estimated peak latencies were shorter and peak amplitudes larger for ABRs from people with normal hearing than those with hearing loss, regardless of method (Table 3). The difference between GMM-FET and visual inspection for each ABR waveform are plotted in reference to the latencies and amplitudes estimated by visual inspection (Fig. 4). A positive bias (solid colored line), or

TABLE 1. Initial values (lower and upper bounds) for parameter *L* corresponding to peak latency

Dataset	Status	SP	Wave I	Wave IV	Wave V	Wave VI
1	NH	0.75 (0.50, 1.00)	1.20 (0.08, 2.00)	4.50 (4.00, 5.00)	5.25 (4.75, 6.25)	6.00 (5.75, 6.50)
	HL	0.75 (0.50, 1.00)	1.50 (1.00, 2.00)	4.50 (4.00, 5.00)	5.25 (4.75, 6.25)	6.00 (5.75, 6.50)
2	NH/Mi	1.00 (0.60, 1.50)	1.50 (1.25, 2.00)	4.25 (3.75, 4.75)	5.25 (4.75, 5.75)	6.25 (6.00, 7.00)

HL, hearing loss; Mi, mild HL; NH, normal hearing; SP, summating potential.

mean difference, indicates the GMM-FET generally estimates larger latencies or amplitudes than visual inspection. The bias in latency between GMM-FET and visual inspection across waveforms was 0.06 to 0.23 msec, depending on wave and hearing status. The GMM-FET bias in amplitude was 0.09 to 0.36 μV. The differences were not normally distributed; therefore, limits of agreement (dashed colored lines) should be interpreted with caution despite Bland-Altman analysis'

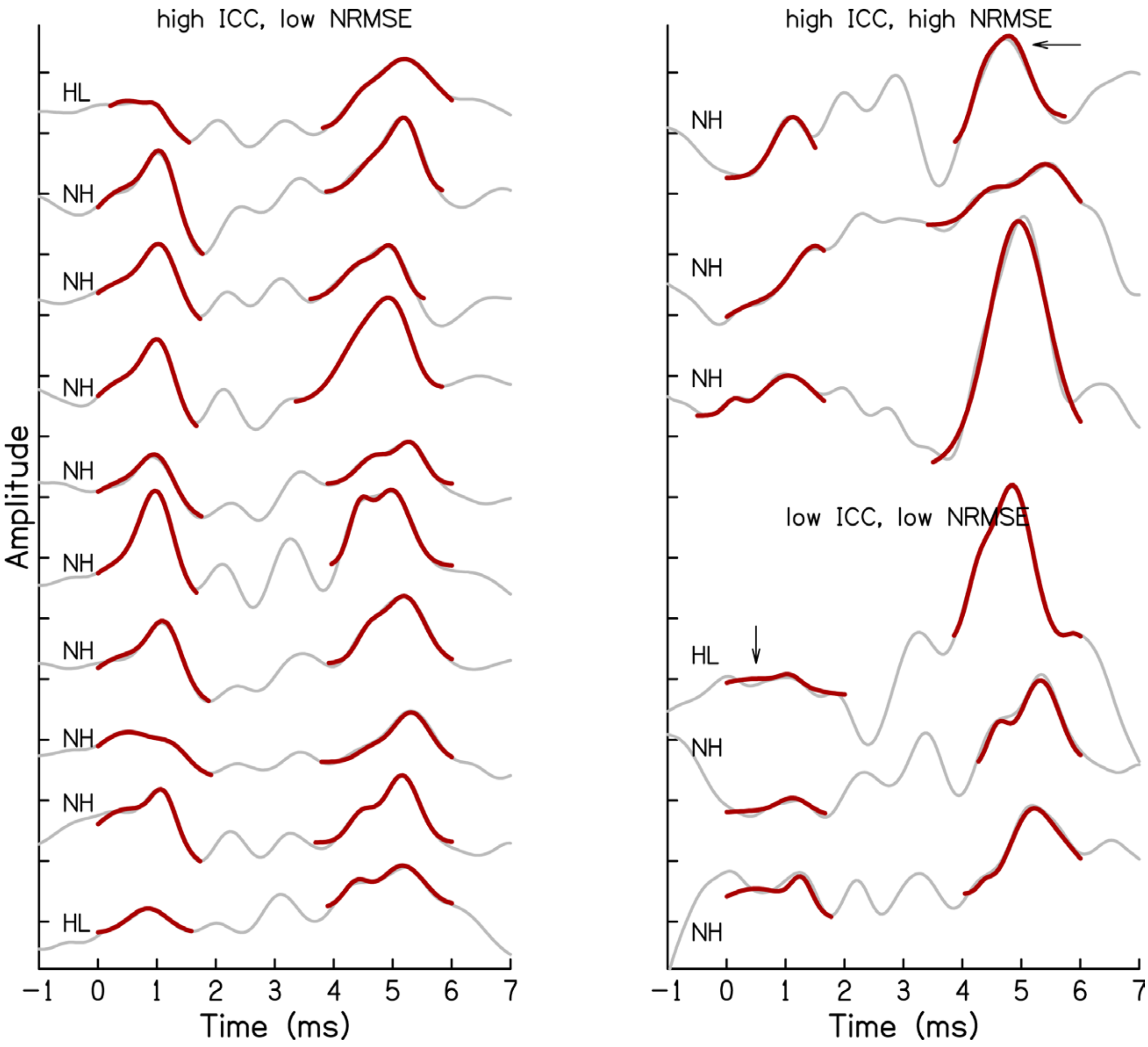

Fig. 2. Sixteen examples of ABR waveforms (gray) from NH and people with hearing loss (HL). The GMM-FET model of the wave I and V complexes (red) is overlayed on the waveform. Left: 10 waveforms with high ICC and low NRMSE, some of the best-fit waveforms. Right: 3 examples of high ICC but relatively high (poor) NRMSE and 3 examples of low NRMSE but low (poor) ICC. Arrows point to exemplar consequences of poor ICC or NRMSE. ABR indicates auditory brainstem response; GMM-FET, Gaussian mixture model-based feature extraction technique; ICC, intraclass correlation; NH, normal hearing; NRMSE, normalized root-mean-square error.

TABLE 2. Goodness-of-fit of GMM-FET for datasets 1 and 2

Dataset	Wave	Status	n	Success Rate (%)	NRMSE % Median (IQR)	ICC Median (IQR)
1	I	NH	99	98	2.9 (3.1)	0.996 (0.010)
		HL	122	96	0.8 (1.6)	0.995 (0.027)
	V	NH	99	100	0.3 (0.4)	0.999 (<0.001)
		HL	122	100	0.4 (0.4)	0.999 (0.002)
2	I	NH/Mi	100	100	1.5 (2.0)	0.996 (0.012)
	V			100	1.4 (1.5)	0.999 (0.003)

GMM-FET, Gaussian mixture model feature extraction technique; HL, hearing loss; ICC, intraclass correlation; IQR, interquartile range; Mi, mild HL; NH, normal hearing; NRMSE, normalized root-mean-square error.

robustness to nonnormal distributions (Bland & Altman 1999). Agreement was also driven, in part, by the amplitude of the wave. Figure 5 plots the same peak latency differences as the *y* axes of Figure 4A, B but as a function of visually estimated peak amplitude (*x* axes of Fig. 4C, D). While there was no correlational trend, a Bruesch-Pagan Test rejected the null hypothesis of homoskedasticity for both wave I ($X^2 = 29.32$ [1], $p < 0.001$) and wave V ($X^2 = 53.08$ [1], $p < 0.001$); there were more and larger latency differences between methods for waves with small peak amplitudes.

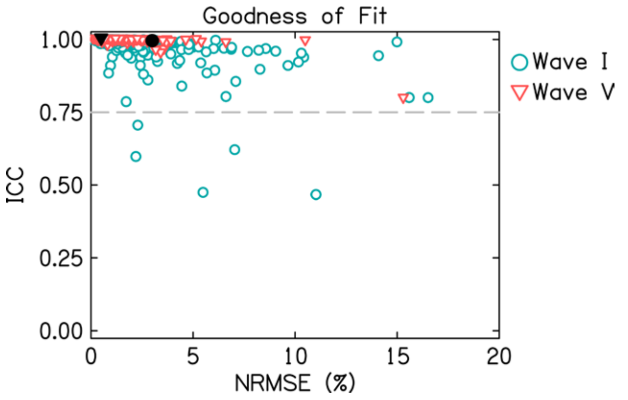

Fig. 3. ICC and NRMSE values for 321 wave I complexes (blue circles) and 321 wave V complexes (red triangles). Median ICC and NRMSE are represented in the filled black symbols, and the cutoff for successful model fit was ICC ≥ 0.75 (dashed gray line). There were seven total wave I complexes that did not meet the cutoff for success (two are not shown because they had an NRMSE > 1). ICC indicates intraclass correlation; NRMSE, normalized root-mean-square error.

DISCUSSION

In this study, we expanded the GMM-FET to include the SP and waves I, IV, V, and VI of the human ABR and assessed the generalizability of the GMM-FET to new datasets. The GMM-FET fit to the waveform was quantified via ICC and NRMSE, and GMM-FET-estimated latencies and amplitudes were compared with those estimated by visual inspection.

Goodness-of-Fit

The success rate of the GMM-FET, determined via ICC ≥ 0.75, was 99%. Visual inspection also did not have a perfect success rate for wave I. This occurs when the peak cannot be visually identified with confidence, as can be the case with noisy waveforms or otherwise poor recordings. In dataset 1, there were three visually unidentifiable waves, resulting in a visual inspection success rate of 99%, the same as the GMM-FET. Goodness-of-fit was comparable to the Kamerer et al. (2020) study, with the majority of waveforms modeled with an ICC > 0.9 and NRMSE < 0.05, or 5% of response amplitude. In dataset 1 in the present study, the wave I complex had worse model fit than wave V, likely because the choice of recording

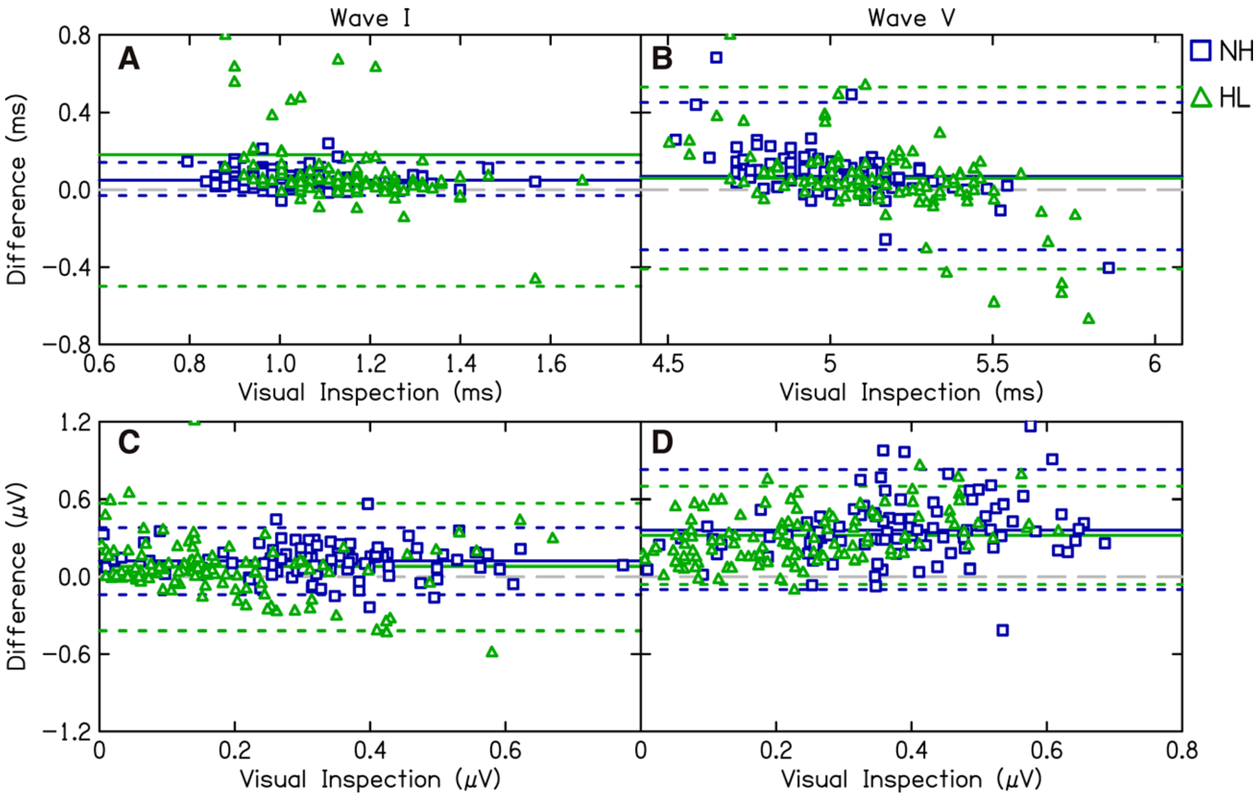

Fig. 4. Difference between modeled and visually inspected wave latencies (A and B) and amplitudes (C and D) for dataset 1. Differences are plotted such that positive values are wave latencies or amplitudes for which the Gaussian mixture estimates a larger value than visual inspection. The agreement is shown separately for waves I (A and C) and wave V (B and D) and for ABR responses from adults with normal hearing (blue squares) and those with hearing loss (green triangles). The gray dashed line at $y = 0$ indicates perfect agreement, colored solid lines indicate bias, and colored dashed lines indicate the upper and lower LoA. ABR indicates auditory brainstem response; HL, hearing loss; LoA, limits of agreement; NH, normal hearing.

parameters for the ABRs used in this analysis were not optimal for recording the SP and wave I. There were not many prominent SPs in the dataset, as evidenced by Figure 2, and many poorer GMM-FET fits to the waveforms (Fig. 2, right panel arrows) were because it was asked to fit an SP that might not be present. Goodman et al. (2023) used the GMM-FET to fit the SP and AP (analogous to wave I of the ABR) of the electrocochleograms in young, normal-hearing adult humans, and found excellent model fit to the data. In contrast with the present study, Goodman et al. used a horizontal electrode montage with tympanic membrane electrodes to improve the signal to noise ratio of the SP and AP (Simpson et al. 2020), whereas the ABRs recorded in the present study used vertex and ear-canal electrodes. In addition, Goodman et al. (2023) used a lower high-pass cutoff frequency than the ABRs from dataset 1, which again would enhance the SP, as it follows the envelope of the organ of Corti response to the stimulus and high level transient stimuli elicit cochlear ringing, dampening the response (Whitfield et al. 1965; Dallos et al. 1972; Ruggero et al. 1997; Zheng et al. 1997, 2011; Durrant et al. 1998; Pappa et al. 2019). Kamerer et al. (2020) also conducted a simulation experiment that found that the GMM-FET was more accurate than visual inspection in estimating the true SP and wave I amplitudes, especially in noisy waveforms and when the SP and wave I overlapped closely in time; however, the simulated SPs had a Gaussian shape and the GMM-FET fit the SP well. In the present study, dataset 2, which had a 100% success rate, recorded with lower high-pass cutoff frequencies than dataset 1. Differences in recording parameters across the various studies might have resulted in better or worse fits to the wave I complex.

Agreement With Visual Inspection

There was overall agreement between GMM-FET and visual inspection because the GMM-FET chose the same peaks to measure as visual inspection. Though despite measuring the same wave, there were some small but notable differences between latencies and amplitudes estimated by the two methods, as shown by deviances from $y = 0$ in Figure 3A, B. The mean peak latencies for both waves I and V were slightly later when estimated by GMM-FET compared with visual inspection, especially earlier peak latencies compared with others in the dataset. There are two explanations for this difference. The first is when there is noise in the waveform or when the waveform is small and near the noise floor, there are often several viable bumps on which to place the peak marker, and in these cases it is not clear which method is picking the "correct" peak. Figure 5 shows the relationship between agreement on peak latency and the amplitude of the wave (per visual inspection). The relationship is heteroskedastic, suggesting latency differences between methods occurred more often and were larger when wave amplitudes were small. The second reason is that the GMM-FET accounts for cumulative effects of preceding

TABLE 3. Agreement between GMM-FET and visual inspection on wave latency and amplitude for dataset 1

Feature	Wave	Status	n	Mean GMM-FET	Mean Visual Insp.	Bias (SD)	LoA	WI-Sub. SD
Latency (msec)	I	NH	96	1.09	1.04	0.06 (0.05)	–0.04, 0.14	0.05
		HL	116	1.32	1.11	0.23 (0.38)	–0.51, 0.97	0.31
	V	NH	98	5.11	5.04	0.07 (0.19)	–0.31, 0.45	0.15
		HL	122	5.20	5.15	0.06 (0.24)	–0.41, 0.53	0.18
Amplitude (μV)	I	NH	96	0.46	0.33	0.12 (0.13)	–0.14, 0.39	0.13
		HL	116	0.27	0.18	0.09 (0.26)	–0.41, 0.60	0.19
	V	NH	98	0.74	0.38	0.36 (0.24)	–0.10, 0.83	0.31
		HL	122	0.55	0.24	0.32 (0.19)	–0.06, 0.70	0.26

GMM-FET, Gaussian mixture model feature extraction technique; HL, hearing loss; LoA, limits of agreement; NH, normal hearing; WI-sub., within subject.

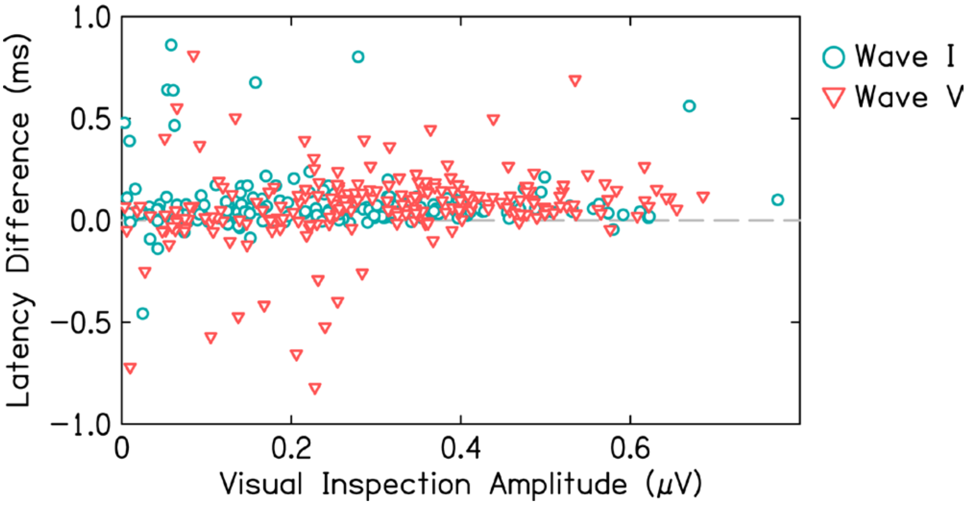

Fig. 5. The difference in estimated peak latency by GMM-FET and visual inspection as a function of peak amplitude estimated by visual inspection. Latency differences are plotted such that positive values are wave latencies for which the Gaussian mixture estimates a larger value than visual inspection. GMM-FET indicates Gaussian mixture model-based feature extraction technique.

waves. For example, in Figure 1, the SP and wave I overlap in time; thus, the amplitude of the offset of the large SP added to the amplitude of the onset of wave I result in a wave I peak that occurs slighter earlier than if the SP did not exist. This also occurs to a lesser degree with the overlap between waves IV and V. Visual inspection can only see the sum of the waves (Fig. 1, black) so the marker is placed at the visual peak (Fig. 1B, black dashed line). Conversely, the GMM-FET extracts wave I or V latency from the individual Gaussian functions (Fig. 1B, green and red, respectively) rather than the summed waveform (gray), and therefore estimates slightly later latencies. If the preceding wave overlaps less, this results in closer estimations between the GMM-FET and visual inspection (e.g., Fig. 1, waves IV and V).

Differences in amplitude estimations between methods are due to two factors. The first is that, again, unlike visual inspection, the GMM-FET accounts for influences of surrounding waves. In Figure 1B example, the GMM-FET estimates a much smaller wave I amplitude than visual inspection due to the large preceding SP. On the other hand, the GMM-FET estimates a larger wave V amplitude than visual inspection in Figure 1 because a visual inspector cannot see the wave V trough and might, depending on the protocol, place the wave V trough marker on the inflection point between waves V and VI (Fig. 1A, “wave V_{trough}”) resulting in an underestimation of wave V amplitude. In our prior study, we conducted a validation experiment that simulated ABR wave I complexes and compared accuracy of the GMM-FET and visual inspection to the true wave amplitudes (Kamerer et al. 2020). The GMM-FET was overall more accurate than visual inspection in estimating wave I. The simulations for which the GMM-FET was more accurate were in noisy waveforms and when waves overlapped closely in time. The second reason for amplitude differences between methods is due to the differences in calculations between visual inspection and the GMM-FET. Visual inspection typically measures peak amplitude from peak to the following trough, whereas the GMM-FET fits Gaussian functions from a baseline. The baseline for this study was set as the amplitude of the first data point in the windowed waveform. For wave I, this was the mean amplitude in a 500 μsec window around time zero, a common measurement of baseline. For wave V, this was the amplitude at $t = 3.5$ msec, around the trough of wave III. This arbitrary baseline for the wave V complex is a limitation of this method in its current state and likely accounts for the larger variability in estimated amplitude differences for wave V than wave I (Fig. 4C, D). Some visual inspection protocols measure wave amplitude from peak to waveform baseline (e.g., the average pre-stimulus waveform amplitude; Mehraei et al. 2016). Because the visual inspection protocol for the study involving dataset 1 measured peak-to-trough, we did not use the waveform baseline to measure wave V but chose a baseline for the entire wave V complex. While the focus of this study was not on the reliability of visual inspection, it is interesting to note that the majority of disagreements between the two visual inspectors in dataset 1 were for the wave V trough latency which is used to calculate wave V amplitude. Of 37 differences that required arbitration by a third visual inspector, there were 33 differences in wave V trough latency, compared with 9 disagreements on wave I peak latency, 11 disagreements on wave I trough latency, and 12 disagreements on wave V peak latency. We argue that these disagreements are a direct result of the overlap between waves V and VI and the peak-to-trough method of amplitude calculation. Amplitude estimations would likely be in better agreement if both the GMM-FET and visual inspectors measured amplitude from the same baseline.

Limitations and Future Directions

One drawback of peak-picking algorithms is that they must be tailored to work on each new dataset, if there are differences in equipment, test parameters, and sample populations. Like peak-picking algorithms, the GMM-FET in its current state requires initial values and bound constraints that must be selected based on the peak latency ranges produced by equipment, chosen test parameters, and population. Table 1 shows slight differences in bound constraints for each dataset, that were recorded using the same equipment, but used different stimuli and filter settings, and included different ranges of hearing sensitivity. To optimize fit for the data, these initial values and bound constraints required manual adjustment. A future direction of the GMM-FET is to create an algorithm used by the GMM to automatically choose initial values and constraints based on test parameters. This would allow the GMM-FET to work on new datasets with minimal input and adjustment.

The GMM-FET had slightly poorer success rate and goodness-of-fit for the wave I complex compared with wave V, likely due to poor fit of the GMM-FET to the SP because the recording parameters for Dataset 1 may have resulted in small or absent SPs. These data suggest the success of the GMM-FET for the SP depends on how large and Gaussian in shape the SP is, which is dependent on the recording parameters such as electrode placement and filtering. The GMM-FET may not always be appropriate for fitting and feature extraction of the SP. Hancock et al. (2021) and Vasilkov et al. (2023) have suggested Fourier analysis as a more appropriate method of separating the SP from wave I of the ABR. The usefulness of the GMM-FET has thus far been assessed for neurodiagnostic ABR procedures that use high presentation levels and click or tone burst stimuli. Neurodiagnostic ABRs typically use high presentation levels to elicit a large response from the auditory system; however, we have only tested the GMM-FET on presentations levels >100 dB peSPL. While the GMM-FET was not created for threshold ABRs, simply because threshold ABR analysis typically involves only determining the presence or absence of a response rather than measurements of wave features, it will be important for future studies to test the limitations of the GMM-FET in terms of how different stimuli and recording parameters affect the usefulness of the technique.

The GMM-FET is a descriptive model of ABR morphology. The benefits of the GMM-FET are that it can extract morphological data of interest near-instantaneously, which could be especially useful in research where hundreds or thousands of ABR waveforms may require analysis for a single study and can extract data that may be difficult to determine visually. Further research is needed to determine whether the additional data provided by the GMM-FET, such as wave width, area, or initial spike (onset) latency, may be useful biomarkers for pathology. In humans, where pathology is difficult or impossible to diagnose in vivo, physiological models of cochlear function (Rønne et al. 2012; Verhulst et al. 2017) could help answer this question.

CONCLUSIONS

A GMM is an appropriate and simple method of automized feature extraction for the ABR. Unlike other template-matching or modeling methods, the GMM is capable of modeling multiple waves of ABR and accounting for interactions between waves that may influence peak latency and amplitude.

ACKNOWLEDGMENTS

The author thank the reviewers for their constructive feedback and ideas that improved both the GMM-FET and the manuscript. The author acknowledges Daniel Rasetshwane, PhD, and Stephen Neely, DSc, for the data used in this study, and Sara Harris, AuD, and Judy Kopun, MS, for their hours of time visually inspecting waveforms and arbitrating differences.

This research was funded by NIDCD 5R01DC016348, NIDCD T32DC000013, and NIGMS P20GM109023 grants.

Data for this analysis were existing data from two prior studies conducted at Boys Town National Research Hospital. All procedures were approved by the Boys Town National Research Hospital Institutional Review Board, and informed consent was obtained from all participants. Participants were paid for their participation. IRB# 16-01-XP.

A. K. is solely responsible for the conception, analysis, interpretation, and writing of this manuscript.

The author has no conflicts of interest to disclose.

Address for correspondence: Aryn M. Kamerer, 1000 Old Main Hl, Logan, UT 84322, USA. E-mail: aryn.kamerer@usu.edu

Received August 8, 2023; accepted February 9, 2024; published online ahead of print February 29, 2024

REFERENCES

Arnold, S. A. (1985). Objective versus visual detection of the auditory brain stem response. *Ear Hear*, *6*, 144–150.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res*, *8*, 135–160.

Bradley, A. P., & Wilson, W. J. (2004). Automated analysis of the auditory brainstem response. Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, ISSNIP, *'04*, 541–545.

Chertoff, M. E. (2004). Analytic treatment of the compound action potential: Estimating the summed post-stimulus time histogram and unit response. *J Acoust Soc Am*, *116*, 3022–3030.

Coats, A. C., & Martin, J. L. (1977). Human auditory nerve action potentials and brain stem evoked responses: Effects of audiogram shape and lesion location. *Arch Otolaryngol*, *103*, 605–622.

Dallos, P., Schoeny, Z. G., Cheatham, M. A. (1972). Cochlear summating potentials: Descriptive aspects. *Acta Otolaryngol*, *302*(Supplement 302), 1–46.

Delgado, R. E., & Ozdamar, O. (1994). Automated auditory brainstem response interpretation. *IEEE Eng Med Biol Mag*, *13*, 227–237.

Don, M., & Elberling, C. (1994). Evaluating residual background noise in human auditory brain-stem responses. *J Acoust Soc Am*, *96*, 2746–2757.

Durrant, J. D., Wang, J., Ding, D., Salvi, R. J. (1998). Are inner or outer hair cells the source of summating potentials recorded from the round window? *J Acoust Soc Am*, *104*, 370–377.

Elberling, C. (1979). Auditory electrophysiology: The use of templates and cross correlation functions in the analysis of brain stem potentials. *Scand Audiol*, *8*, 187–190.

Galambos, R., & Hecox, K. E. (1978). Clinical applications of the auditory brainstem response. *Otolaryngol Clin North Am*, *11*, 709–722.

Goldstein, M. H., & Kiang, N. Y. S. (1958). Synchrony of neural activity in electric responses evoked by transient acoustic stimuli. *J Acoust Soc Am*, *30*, 107–114.

Goodman, S. S., Lichtenhan, J. T., Jennings, S. G. (2023). Minimum detectable differences in electrocochleography measurements: Bayesian-based predictions. *J Assoc Res Otolaryngol*, *24*, 217–237.

Gorga, M. P., Worthington, D. W., Reiland, J. K., Beauchaine, K. A., Goldgar, D. E. (1985). Some comparisons between auditory brain stem response thresholds, latencies, and the pure-tone audiogram. *Ear Hear*, *6*, 105–112.

Hancock, K. E., O'Brien, B., Santarelli, R., Liberman, M. C., & Maison, S. F. (2021). The summating potential in human electrocochleography: Gaussian models and Fourier analysis. *J Acoust Soc Am*, *150*, 2492–2502. https://doi.org/10.1121/10.0006572

Irino, T., & Patterson, R. D. (2009). A dynamic compressive gammachirp auditory filterbank. *IEEE Trans Audio Speech Lang Process*, *14*, 2222–2232.

Jerger, J., & Mauldin, L. (1978). Prediction of sensorineural hearing level from the brain stem evoked response. *Arch Otolaryngol*, *104*, 456–461.

Kamerer, A. M., Kopun, J. G., Fultz, S. E., Allen, C., Neely, S. T., & Rasetshwane, D. M. (2019a). Examining physiological and perceptual consequences of noise exposure. *J Acoust Soc Am*, *146*, 3947–3959. https://doi.org/10.1121/1.5132291

Kamerer, A. M., Aubuchon, A., Fultz, S. E., Kopun, J. G., Neely, S. T., & Rasetshwane, D. M. (2019b). The role of cognition in common measures of peripheral synaptopathy and hidden hearing loss. *Am J Audiol*, *28*, 843–856. https://doi.org/10.1044/2019_AJA-19-0063

Kamerer, A. M., Neely, S. T., Rasetshwane, D. M. D. M. (2020). A model of auditory brainstem response wave I morphology. *J Acoust Soc Am*, *147*, 25. 31.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*, *15*, 155–163.

Krumbholz, K., Hardy, A. J., de Boer, J. (2020). Automated extraction of auditory brainstem response latencies and amplitudes by means of non-linear curve registration. *Comput Methods Programs Biomed*, *196*, 105595.

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

Melcher, J. R., & Kiang, N. Y. (1996). Generators of the brainstem auditory evoked potential in cat III: identified cell populations. *Hear Res*, *93*, 52–71.

Melcher, J. R., Guinan, J. J. Jr, Knudson, I. M., Kiang, N. Y. (1996a). Generators of the brainstem auditory evoked potential in cat. II. Correlating lesion sites with waveform changes. *Hear Res*, *93*, 28–51.

Melcher, J. R., Knudson, I. M., Fullerton, B. C., Guinan, J. J., Norris, B. E., Kiang, N. Y. S. (1996b). Generators of the brainstem auditory evoked potential in cat. I. An experimental approach to their identification. *Hear Res*, *93*, 1–27.

Pappa, A. K., Hutson, K. A., Scott, W. C., Wilson, J. D., Fox, K. E., Masood, M. M., Giardina, C. K., Pulver, S. H., Grana, G. D., Askew, C., Fitzpatrick, D. C. (2019). Hair cell and neural contributions to the cochlear summating potential. *J Neurophysiol*, *121*, 2163–2180.

Picton, T., Hunt, M., Mowrey, R., Rodriguez, R., Maru, J. (1988). Evaluation of brain-stem auditory evoked potentials using dynamic time warping. *Electroencephalogr Clin Neurophysiol*, *71*, 212–225.

Rønne, F. M., Dau, T., Harte, J. M., Elberling, C. (2012). Modeling auditory evoked brainstem responses to transient stimuli. *J Acoust Soc Am*, *131*, 3303–3313.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., Robles, L. (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J Acoust Soc Am*, *101*, 2151–2163.

Salarian, A. (2016). Intraclass Correlation Coefficient (ICC) Arash Salarian (2023). *Intraclass Correlation Coefficient (ICC)*. https://www.mathworks.com/matlabcentral/fileexchange/22099-intraclass-correlation-coefficient-icc. (1.3.1.0)

Simpson, M. J., Jennings, S. G., Margolis, R. H. (2020). Techniques for obtaining high-quality recordings in electrocochleography. *Front Syst Neurosci*, *14*, 1–12.

Sininger, Y. S. (1993). Auditory brain stem response for objective measures of hearing. *Ear Hear*, *14*, 23–30.

Suthakar, K., & Liberman, M. C. (2019). A simple algorithm for objective threshold determination of auditory brainstem responses. *Hear Res*, *381*, 107782.

Valderrama, J. T., De la Torre, A., Alvarez, I., Segura, J. C., Thornton, A. R. D., Sainz, M., Vargas, J. L. (2014). Automatic quality assessment and peak identification of auditory brainstem responses with fitted parametric peaks. *Comput Methods Programs Biomed*, *114*, 262–275.

Vannier, E., Adam, O., Motsch, J. -F. (2002). Objective detection of brainstem auditory evoked potentials with a priori information from higher presentation levels. *Artif Intell Med*, *25*, 283–301.

Vasilkov, V., Liberman, M. C., Maison, S. F. (2023). Isolating auditory-nerve contributions to electrocochleography by high-pass filtering: A better biomarker for cochlear nerve degeneration? *JASA Express Lett*, *3*, 024401.

Verhulst, S., Altoè, A., Vasilkov, V. (2017). Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hear Res*, *360*, 55–75.

Vidler, M., & Parker, D. (2004). Auditory brainstem response threshold estimation: Subjective threshold estimation by experienced clinicians in a computer simulation of the clinical test. *Int J Audiol*, *43*, 417–429.

Whitfield, I. C., Ross, H. F., Wakefield, I. C., Ross, H. F. (1965). Cochlear-microphonic and summating potentials and the outputs of individual hair-cell generators. *J Acoust Soc Am*, *38*, 126–131.

Zheng, X. Y., Ding, D., McFadden, S. L., Henderson, D. (1997). Evidence that inner hair cells are the major source of cochlear summating potentials. *Hear Res*, *113*, 76–88.

Zheng, J., Ramamoorthy, S., Ren, T., He, W., Zha, D., Chen, F., Magnusson, A., Nuttall, A. L., Fridberger, A. (2011). Persistence of past stimulations: Storing sounds within the inner ear. *Biophys J*, *100*, 1627–1634.