Effects of Artifact Rejection and Bayesian Weighting on the Auditory Brainstem Response During Quiet and Active Behavioral Conditions

Jason Tait Sanchez
Donald Gans
Kent State University,
Northeastern Ohio Universities College of Medicine,
Rootstown, OH

Purpose: To evaluate the effects of 2 noise reduction techniques on the auditory brainstem response (ABR).

Method: ABRs of 20 normal hearing adults were recorded during quiet and active behavioral conditions using 2 stimulus intensity levels. Wave V amplitudes and residual noise root-mean-square values were measured following the offline application of artifact rejection and Bayesian weighting. Repeated measures analysis of variance and Bonferroni adjusted pairwise *t* tests were utilized to evaluate significant main effects and interactions between the 2 noise reduction techniques.

Results: ABRs recorded during the quiet behavioral condition resulted in minimal differences in wave V amplitude and noise reduction improvement, suggesting that the 2 techniques were equally effective under ideal recording situations. During the active behavioral condition, however, the techniques differed significantly in the ability to preserve the evoked potential and reduce noise. Consequently, strict artifact rejection levels resulted in an inherent underestimation of wave V amplitudes when compared with the Bayesian approach.

Conclusion: Artifact rejection had a detrimental effect on waveform morphology of the ABR. This could lead to difficulty in ABR interpretation when patients are active and ultimately result in diagnostic errors.

Key Words: artifact rejection, evoked potential, Bayesian weighting, auditory brainstem response

Amplitude measurements of the auditory brainstem response (ABR) provide valuable information regarding the peripheral auditory pathway and lower brainstem nuclei (Boston & Møller, 1984; Chandrasekhar, Brackmann, & Devgan, 1995; Coats & Martin, 1978; Don, Masuda, Nelson, & Brackmann, 1997; Kotlarz, Eby, & Borton, 1992; Marangos, Maier, Merz, & Laszig, 2001; Wilson, Hodgson, & Gustafson, 1992). This important measurement is negatively affected by excessive noise, and researchers have continuously made efforts to reduce such contaminants by implementing noise reduction techniques on the averaged ABR (Don & Elberling, 1994; Kavanagh & Franks, 1989; Scherg & Von Cramon, 1984; Turetsky, Raz, & Fein, 1988).

In general, techniques such as filtering, signal averaging, and artifact rejection have been employed to facilitate the extraction of the evoked potential (EP) from unwanted noise (Kavanagh & Franks, 1989; Schimmel, 1967). An underlying assumption made when utilizing such techniques is that the signal of interest (i.e., the EP) is preserved while noise is substantially reduced, thus improving the signal-to-noise ratio (SNR). Filtering, however, provides minimal improvement of the SNR because the frequency spectrum of noise often overlaps with the frequency composition of the EP (Boston & Ainslie, 1980; Elton, Scherg, & Von Cramon, 1984; Marsh, 1988; Osterhammel, 1981). Similarly, signal averaging theoretically reduces noise by the square root of the number of sweeps in the averaged response. This theoretical assumption, however, is not always attainable due to random noise variations caused by episodic movement (Don, Elberling, & Waring, 1984). Artifact rejection, on the other hand, evaluates the amplitude of the incoming noise from the electrodes for individual sweeps. If the noise exceeds a predetermined microvolt level, the sweep is rejected from the computer memory

1059-0889/06/1502-0154

and not included in the averaging process. Since the EP is considered a deterministic component (i.e., invariant in time and amplitude), it is generally assumed that its amplitude is minimally affected while noise is reduced by the square root of the number of sweeps in the averaged response. Thus, the averaged resulting trace better reflects the true EP when compared with individual sweeps.

Contrary to such assumptions, artifact rejection can have a detrimental effect on ABR testing, which has warranted the development of superior noise reduction techniques (Don & Elberling, 1994; Stecker, 2002). Hoke, Ross, Wickesberg, and Lütkenhöner (1984) used a weighted averaging technique that was more efficient at estimating the EP from noise when compared with the traditional averaging technique. Numerous others have incorporated weighted averaging in EP testing, and there is a general consensus among these studies that such a technique is efficient in reducing excessive noise (Bezerianos, Laskaris, Fotopoulos, & Papathahasopoulos, 1995; Davila & Mobin, 1992; Don et al., 1984; Elberling & Don, 1984; Elberling & Wahlgreen, 1985; Gasser, Mocks, & Verleger, 1983; Gerull, Graffunder, & Wernicke, 1996; Hoke et al., 1984; John, Dimitrijevic, & Picton, 2001; Lütkenhöner, Hoke, & Pantev, 1985; Sparacino, Milani, Arslan, & Cobelli, 2002; Wicke, Gogg, Wallace, & Allison, 1978; Wong & Bickford, 1980).

Bayesian weighting uses an estimating technique (Elberling & Don, 1984) to reduce destructive effects of noise variation on the ABR (Elberling & Wahlgreen, 1985). The Bayesian approach to weighted averaging is a variation on the traditional averaging technique and is derived from a statistical method known as Bayesian inference. The approach stems from condition probability, which uses a mathematical model that is contingent upon several theoretical assumptions. Specifically, Bayesian inference is established on three principles: a priori knowledge, the likelihood function, and a posteriori information. Bayesian inference relies on the principle that adding new data through the likelihood function to established a priori knowledge, updated a posteriori information will be produced (Elberling & Wahlgreen, 1985). Such principles are well suited for ABR testing, because new data are continuously added to prior data and averaged.

Few, if any, studies have directly evaluated whether artifact rejection has a destructive effect on waveform morphology compared with Bayesian weighting. Don and Elberling (1994) compared the two techniques and found the Bayesian approach to be the superior technique for improving the SNR. They also suggested that differences in SNRs were evident when patients presented with episodic noise. They, however, did not compare the two techniques during systematic active behavioral conditions, nor did the study evaluate the effects of the techniques on waveform morphology, in particular, on the peak-to-trough amplitude of wave V. Therefore, the major goal of this study was to demonstrate how different methods of reducing noise in the averaged ABR affect the amplitude measurement of wave V.

Method

Participants

Twenty normal hearing adults (15 women, 5 men) were randomly selected from the Kent State University Speech and Hearing program. Each participant signed a consent form approved by the Human Subject Research Review Board at Kent State University prior to testing. The mean age of the participants was 23 years (*SD* = 5 years). Otoscopic examinations and tympanometry were performed to rule out conductive problems that might preclude audiometric and ABR testing. Pure-tone audiometry was performed with a Grason-Stadler GSI-61 audiometer using Etymotic Research ER-3A insert earphones. All participants had pure-tone thresholds less than or equal to 10 dB HL (American National Standards Institute, 1996; Carhart & Jerger, 1959) for octave frequencies ranging from 250 to 8000 Hz at the time of ABR data collection.

Apparatus

ABRs were recorded differentially using a silver-silver chloride disk electrode applied to the high forehead (active) and disposable gold-foil tiptrodes applied to the ipsilateral (reference) and contralateral (ground) ear canals. Overall electrode impedances were less than 5.0 kΩ, and interelectrode impedances were less than 1.0 kΩ for each participant. Scalp activity was amplified 10^4 using two CWE differential amplifiers (BMA 831, BMA 830) and analog filtered between 100 Hz and 5000 Hz at 12 dB per octave slope. A second offline digital filter was implemented at a bandpass setting between 100 and 3000 Hz.

A 100-μs rectangular voltage pulse was presented to a Coulbourn audio-mixer amplifier and adjusted to 1 V peak-to-peak. Shielded ER-3A insert earphones served as transducers. Rarefaction clicks were presented 25.1 per second at 104 and 74-dB peak-to-peak equivalent sound pressure level (p-pe SPL) with a 1000-Hz tone as reference. The 104-dB p-pe SPL click was 60 dB above the average perceptual detection threshold, or 60 dB nHL, while the 74-dB p-pe SPL click was 30 dB nHL. Average thresholds using clicks were evaluated in 1-dB steps for 10 normal hearing participants to determine 0 dB nHL. The rationale for using two intensity levels was to mimic a neurodiagnostic suprathreshold technique (60 dB nHL) and a near-threshold estimation procedure (30 dB nHL). The ER-3A insert earphones were calibrated using a Bruel & Kjaer 1613 sound level meter fitted with a 2 cm^3 coupler.

Custom computer software was developed for data recordings and analysis. Data were sampled at a rate of 48 kHz for 15 ms poststimulus onset (sweep) using a Quantum DSP-3210 analog-to-digital converter board connected to a Dell Pentium IV personal computer. Each sweep was composed of 720 digitized points and saved to disk without any modifications. Sampled sweeps were analyzed offline for the first 7.5 ms of the 15 ms sweep. Electrode activity was amplified such that the maximum clipping level was set to an effective value of ±240 μV or ±2.4 V input to the analog-to-digital converter. In no case was this clipping level reached for any participant during any test condition.

Procedures

ABRs were recorded from each participant during two conditions (i.e., quiet and active) at two intensity levels

(60 and 30 dB nHL). Therefore, a total of four recordings were stored on disk. Each recording contained 16,384 sweeps that were later analyzed offline and initially reprocessed to form three sets of averages based on individual noise reduction techniques: Bayesian weighted average, artifact rejection equal noise (AR_{EN}) average, and artifact rejection ±10 μV (AR_{10}) average. This resulted in 12 traces for each participant or a total of 240 ABR traces for 20 normal hearing adults.

Quiet condition ABR recordings were performed in the traditional manner. Participants were placed in a reclining chair and asked to remain in a quiet and relaxed state. For the active condition, participants were instructed to perform an activity that introduced excessive physiological movement. A computer program was used to instruct participants to randomly perform one of three tasks periodically throughout the active ABR recording. Tasks included opening and closing the mouth, moving the head side to side, and moving the head up and down. The computer used to generate the animated program was placed in an adjacent sound suite to reduce electrical interference. The viewing monitor was placed approximately 1 m in front of the participant. This active condition was used to simulate difficult-to-test populations, such as infants or children, who often show high levels of periodic movement.

Each of the three noise reduction techniques were implemented offline, and the resulting ABR traces were subsequently analyzed. For Bayesian weighting, the residual noise was first estimated according to the variance approach described by Elberling and Don (1984). The noise was estimated for a block of 256 sweeps by computing the sweep-to-sweep variance of a single time point in the sweep. The single time point 6 ms after stimulus onset, corresponding to the 288th digitized point, was utilized. Thus, for a block of 256 sweeps, 256 discrete single point values were used in computing the variance and estimating the noise. Blocks of sweeps were then weighted inversely proportional to the amount of noise estimated for that particular block (Elberling & Wahlgreen, 1985). When the estimated noise was large, that particular block received proportionally less weight in the final average. For the current study, two weighted averages were formed using blocks of 16 (i.e., 4,096 sweeps) and 64 (16,384 sweeps). A detailed description of the averaging technique is found in the Appendix.

While Elberling and Don (1984) used a calculation related to the single point variance as an estimate of residual noise in the averaged response, herein the residual noise root-mean-square (RMS) values for both Bayesian weighting and artifact rejection following the theoretical removal of the EP from the ABR recording were calculated. This was achieved by storing 256 consecutive sweeps in alternate buffers. Buffers were subtracted to obtain an overall noise RMS value. This process ensured that any deterministic elements of the recording (i.e., the EP) were eliminated (John et al., 2001; Schimmel, 1967). However, it should be noted that the subtraction method increased noise levels relative to each individual buffer. To correct for this, buffers were averaged to reduce further residual noise and to obtain accurate noise RMS levels. The first 4,096 accepted sweeps were used in estimating noise RMS levels for Bayesian weighting and artifact rejection.

For artifact rejection, two rejection levels were used: AR_{EN} and AR_{10}. If any data point exceeded the rejection level, the entire sweep was rejected and not included in the averaging process. For $AR_{EN,}$ the rejection levels were systematically reduced in 1-μV steps until the noise RMS values were equal to the noise RMS values attained by Bayesian weighting (16 blocks). Mean artifact rejection levels for AR_{EN} were 26 μV (*SD* = 19 μV) and 43 μV (*SD* = 26 μV) during the quiet and active ABR conditions, respectively. This method was used to compare Bayesian weighting and AR_{EN} based on identical noise RMS values. For AR_{10}, a fixed artifact rejection level of ±10 μV was used. The first 4,096 accepted sweeps were used for each artifact rejection criteria.

Bayesian weighting and artifact rejection calculations (i.e., wave V amplitude measurements and noise RMS values) were performed using custom software without experimenter intervention. The amplitude of wave V was measured from the first positive peak of the largest waveform component 5 ms poststimulus onset to the most negative following trough 2 ms before positive deflection. If wave V appeared trough-like, round or bimodal, the last point before rapid negative reflection was identified as the peak (Durieux-Smith, Edwards, Picton, & MacMurray, 1985; Stuart & Yang, 1994). The second author and one independent scorer who was unaware of the purpose of the study and blind to the test conditions subjectively selected the peak-to-trough amplitude of wave V for 50% of the data. An interjudge agreement coefficient (Pearson *r*) was .93 for wave V amplitude measurement.

Results

Wave V Amplitude Analysis

Table 1 summarizes the descriptive statistics for mean wave V amplitudes as a function of stimulus level, ABR test condition, and noise reduction technique. A three-factor repeated measures analysis of variance (ANOVA) was performed to study differences in mean wave V amplitudes as a

Table 1. Mean wave V amplitude measurement in microvolts (μV) and standard deviations of the mean as a function of stimulus intensity level, auditory brainstem response (ABR) test condition, and noise reduction technique (NRT).

Stimulus intensity level	ABR test condition	NRT	*M*	*SD*
60 dB nHL	Quiet	Bayesian	0.357	0.114
		AR_{EN}	0.322	0.098
		AR_{10}	0.317	0.095
	Active	Bayesian	0.368	0.112
		AR_{EN}	0.288	0.106
		AR_{10}	0.182	0.084
30 dB nHL	Quiet	Bayesian	0.181	0.062
		AR_{EN}	0.154	0.053
		AR_{10}	0.152	0.060
	Active	Bayesian	0.220	0.086
		AR_{EN}	0.156	0.076
		AR_{10}	0.104	0.058

Note. AR_{EN} = artifact rejection equal noise; AR_{10} = artifact rejection ±10 μV; *n* = 20.

Table 2. Repeated measures analysis of variance investigating wave V amplitude as a function of NRT, ABR test condition, and stimulus intensity level.

Source	*F*	*df*	*p*	η^2	φ
NRT	36.90	2	<.0001	.660	1.000
ABR test condition	4.62	1	.05	.196	0.271
Stimulus intensity level	120.12	1	<.0001	.863	1.000
NRT × Condition	27.67	2	<.0001	.593	1.000
NRT × Intensity	3.38	2	.05	.151	0.340
Condition × Intensity	5.44	1	.03	.222	0.328
NRT × Condition × Intensity	2.27	2	.07	.126	0.261

Note. Statistical analysis was performed with a significance alpha level of .05.

function of ABR test condition, stimulus intensity level, and noise reduction technique. Statistical analyses were performed with a significance alpha level of $p < .05$.

The results of the ANOVA are presented in Table 2. Statistically significant main effects were found for noise reduction techniques and stimulus intensity level. The intensity effect was expected (i.e., wave V amplitude being larger for 60 dB nHL vs. 30 dB nHL) and therefore not evaluated due to the well-documented aspects of stimulus level on wave V amplitude (Hall, 1992; Jacobson, 1994). A statistically significant interaction between noise reduction techniques and ABR test conditions was also observed (see Table 2). Post hoc mean pairwise comparisons were performed using Bonferroni adjusted *t* tests ($p < .01$; Keselman, Keselman, & Shaffer, 1991; Sato, 1996). Figure 1 shows mean wave V amplitudes recorded during quiet and active ABR conditions for each noise reduction technique. During the quiet ABR condition, there were no significant wave V amplitude differences between the techniques (see Figure 1a). During the active ABR condition, however, AR_{10} resulted in significantly smaller wave V amplitudes than AR_{EN}, $t(19) = 4.85$, $p < .01$, and Bayesian weighting, $t(19) = 8.22$, $p < .01$ (see Figure 1b). In

Figure 1. Mean wave V amplitudes as a function of noise reduction techniques recorded during the quiet (a) and active (b) auditory brainstem response (ABR) conditions. BW = Bayesian weighting; AR_{EN} = artifact rejection equal noise; AR_{10} = artifact rejection ±10 μV. Error bars represent 1 *SEM*. Asterisks identify statistically significant Bonferroni adjusted pairwise comparisons (p < .01).

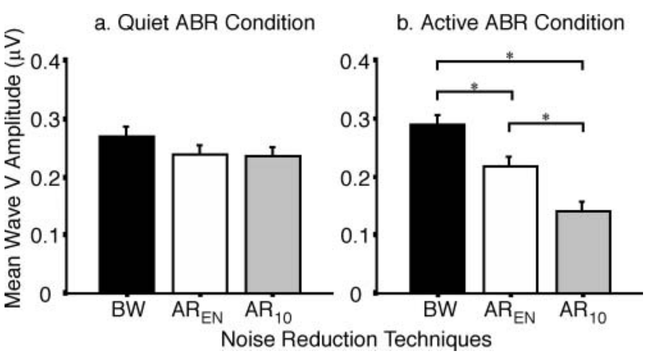

Table 3. Mean noise root-mean-square (RMS) values and standard deviations of the mean as a function of stimulus intensity level, ABR test condition, and NRT.

Stimulus intensity level	ABR test condition	NRT	*M*	*SD*
60 dB nHL	Quiet	Bayesian	20.7	6.8
		AR_{EN}	20.6	6.8
		AR_{10}	19.3	4.0
	Active	Bayesian	66.7	28.5
		AR_{EN}	66.6	28.3
		AR_{10}	21.3	4.6
30 dB nHL	Quiet	Bayesian	23.9	11.4
		AR_{EN}	23.8	11.5
		AR_{10}	18.4	6.1
	Active	Bayesian	52.9	27.0
		AR_{EN}	53.2	27.1
		AR_{10}	22.2	11.2

Note. $n = 20$.

addition, AR_{EN} resulted in significantly smaller wave V amplitudes than Bayesian weighting, $t(19) = 4.68$, $p < .01$. Overall, artifact rejection yielded significantly smaller wave V amplitudes in 90% of the traces (108 of 120 traces) compared with Bayesian weighting. Thus, nonstationary noise plays a role in how effective the techniques are in extracting the EP.

Noise RMS Analysis

Table 3 summarizes the descriptive statistics for mean noise RMS values as a function of stimulus level, ABR test condition, and noise reduction technique. Similar to the wave V amplitude analysis, a three-factor repeated measures ANOVA was undertaken to investigate differences in mean noise RMS values as a function of ABR test condition, stimulus intensity level, and noise reduction technique. The results of the ANOVA are presented in Table 4.

Statistically significant main effects were found for noise reduction techniques and ABR test condition. The test condition effect was expected. That is, physiological movement produced during the active condition resulted in significantly more noise than the quiet condition. A statistically significant interaction between noise reduction techniques and ABR test conditions was also observed (see Table 4).

Table 4. Repeated measures analysis of variance investigating noise RMS values as a function of NRT, ABR test condition, and stimulus intensity level.

Source	*F*	*df*	*p*	η^2	φ
NRT	50.65	2	<.0001	.727	1.000
ABR test condition	58.98	1	<.0001	.756	1.000
Stimulus intensity level	1.80	1	.20	.087	0.089
NRT × Condition	42.70	2	<.0001	.692	1.000
NRT × Intensity	1.66	2	.20	.081	0.137
Condition × Intensity	3.88	1	.06	.170	0.219
NRT × Condition × Intensity	5.41	2	.03	.222	0.589

Note. Statistical analysis was performed with a significance alpha level of .05.

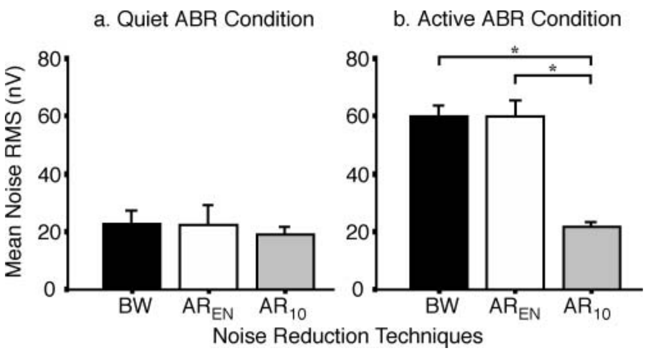

Figure 2. Mean noise root-mean-square (RMS) values as a function of noise reduction techniques recorded during the quiet (a) and active (b) ABR conditions. Error bars represent 1 *SEM*. Asterisks identify statistically significant Bonferroni adjusted pairwise comparisons ($p < .01$).

Figure 2 shows mean noise RMS values achieved by each technique during quiet and active ABR conditions. Post hoc mean pairwise comparisons were performed using Bonferroni adjusted *t* tests ($p < .01$). During the quiet ABR condition, there were no significant noise RMS differences between the noise reduction techniques (see Figure 2a). During the active condition, however, AR_{10} resulted in significantly less noise than AR_{EN}, $t(19) = 7.13$, $p < .01$, and Bayesian weighting, $t(19) = 7.12$, $p < .01$ (see Figure 2b). There were no significant differences between Bayesian weighting and AR_{EN}, $t(19) = 0.22$, $p = .82$, which supports the conclusion that noise levels obtained by AR_{EN} were equal to Bayesian weighting. The fact that AR_{10} reduced more noise than the other two techniques when patients were active again leads one to suggest that nonstationary noise plays a role in technique effectiveness.

One could further suggest that the two techniques work on different aspects of the ABR, prompting the following question: What effect would AR_{10} and Bayesian weighting have on the amplitude of wave V if noise RMS levels were made equivalent? To address this question, we empirically investigated Bayesian weighting and AR_{10} based on equal amounts of noise reduction. To optimize the noise RMS for Bayesian weighting, all 64 blocks of sweeps (i.e., 16,384 sweeps) were used. It should be noted that no sweeps were clipped by the input amplifier. Therefore, all 16,384 sweeps were included in the overall averaged response. This technique was called Bayesian weighting 16k (BW_{16k}). The residual noise levels achieved by BW_{16k} were not target noise levels. Nonetheless, they were directly compared with AR_{10} in order to evaluate our attempt at achieving equal noise levels between the two techniques.

Figure 3 shows mean wave V amplitudes and noise RMS values achieved by BW_{16k} and AR_{10}. Paired samples *t* tests revealed that AR_{10} resulted in significantly smaller wave V amplitudes than BW_{16k}, $t(19) = 6.23$, $p < .05$ (see Figure 3a). However, there was not a significant difference in noise reduction between the two techniques, $t(19) = 1.09$, $p = .19$, which demonstrates that our attempt to achieve equal noise RMS was successful (see Figure 3b). These results further

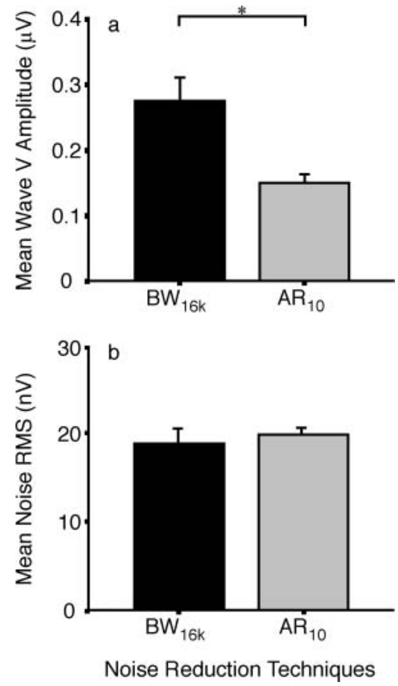

Figure 3. Mean wave V amplitudes (a) and noise RMS values (b) as a function of noise reduction techniques. BW_{16k} = Bayesian weighting averaged for 16,384 sweeps. Error bars represent 1 *SEM*. Asterisks identify statistically significant paired sample comparisons ($p < .05$).

suggest that noise reduction achieved by BW_{16k} did not affect EP amplitude, whereas noise reduction achieved by AR_{10} resulted in reduced EP amplitudes.

Figure 4 best illustrates the main results of this study. Grand averaged ABR traces recorded for a participant during the

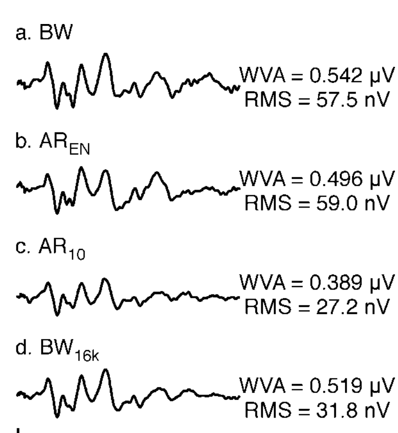

Figure 4. Grand averaged ABR traces for a participant recorded during the active condition at 60 dB nHL. Raw data were reprocessed to form 4 sets of averages based on individual noise reduction techniques: (a) BW, (b) AR_{EN}, (c) AR_{10}, and (d) BW_{16k}. WVA = Wave V amplitude; vertical bar = 0.25 μV; horizontal bar = 1 ms.

active condition at 60 dB nHL are shown. Raw data were reprocessed to form four sets of averages based on individual noise reduction techniques: Bayesian weighting, AR_{EN}, AR_{10}, and BW_{16k}. Using 4,096 sweeps, AR_{10} resulted in a reduced wave V amplitude (0.389 μv) compared with AR_{EN} (0.496 μV) and Bayesian weighting (0.542 μV). Additionally, AR_{10} resulted in more noise reduction (27.2 nV) than Bayesian weighing (57.5 nV) and AR_{EN} (59.0 nV). It should be noted, however, that before attaining 4,096 accepted sweeps, the total number of rejected sweeps for AR_{EN} and AR_{10} was 278 and 1,933, respectively. When Bayesian weighting noise level was equated to AR_{10} (AR_{10} = 27.2 nV vs. BW_{16k} = 31.8 nV), the amplitude of wave V for AR_{10} was significantly reduced compared with BW_{16k} (BW_{16k} = 0.519 μV vs. AR_{10} = 0.389 μV). This result further supports our conclusion that artifact rejection has a destructive effect on the amplitude of wave V despite a substantial reduction in noise.

Discussion

Bayesian weighting and artifact rejection are well-established techniques that reduce noise during ABR testing, thus improving the SNR (Don & Elberling, 1994; Elberling & Wahlgreen, 1985; Pantev & Khvoles, 1984). This is especially true when participants are quiet and relaxed during testing. Consequently, one would expect that the two techniques are equally effective during ideal testing conditions. The results from the current study support this expectation. When participants were quiet and relaxed, the differences in wave V amplitudes and noise RMS values were minimal, suggesting that the techniques are equally effective in ABR testing under ideal situations. To our knowledge, however, no study has investigated systematically whether the two techniques are equally effective when patients are periodically active. We found that during the active behavioral condition, the two techniques differed significantly in their ability to extract the EP and in their ability to reduce noise. Thus, the findings of this study are twofold. First, strict artifact rejection levels have a detrimental effect on the amplitude measurement of wave V. Second, setting strict levels significantly reduces the noise but does not always guarantee an improvement in waveform morphology. This is especially evident when participants are periodically active.

The initial use of 4,096 averaged sweeps allowed us to compare Bayesian weighting with artifact rejection based on equivalent noise levels. That is, we were able to compare the two techniques based on equal noise by systematically readjusting the rejection criterion until the residual noise achieved by artifact rejection was equivalent to Bayesian weighting. During the active condition ABR recording, the average artifact rejection level used to obtain equal noise (i.e., AR_{EN}) with Bayesian weighting was 43 μV, which is a lenient rejection level but not unusual for clinical use (Schwartz & Schwartz, 1991). As defined and empirically determined, the amount of noise reduction achieved by the two techniques was equal, but unexpectedly, the amplitude of wave V was significantly more reduced for AR_{EN} than for Bayesian weighting. Thus, we concluded that artifact rejection has a destructive effect on ABR testing during active behavioral conditions.

Further support of the above conclusion was found when the artifact rejection criterion was made strict (i.e., 10 μV) but still within levels that are used clinically. Again, the noise in the traces for all participants was significantly reduced from the less strict criterion of AR_{EN}. Nevertheless, and more surprisingly, the wave V amplitude was also significantly reduced. It should be noted, however, that residual noise in the averaged trace contributes to the amplitude of wave V. In theory, the larger the residual noise, the greater its contribution to the peak-to-trough amplitude of wave V. Thus, the smaller wave V amplitudes, as seen with AR_{10}, could also be due to less residual noise in the averaged trace. To address the issue, we compared AR_{10} with Bayesian weighting based on equal noise (i.e., BW_{16k}). Surprisingly, the amplitude of wave V was significantly reduced for AR_{10} compared with BW_{16k} despite the fact that both techniques essentially reduced the same amounts of noise.

Such results raise an obvious question: How does artifact rejection reduce an already small EP that is embedded in large amounts of noise? It is generally assumed that the EP is deterministic (i.e., invariant in time and amplitude) and therefore unaffected by noise reduction techniques. If this assumption is true, techniques that reduce equal amounts of noise should yield similar wave V amplitudes. In the current study, this assumption was true for Bayesian weighting but not for artifact rejection.

One possible explanation is that the EP is not deterministic and that the rejected sweeps contain a different subgroup of EPs than accepted sweeps. This explanation is not likely because the Bayesian approach sampled all sweeps and showed no such reduction. A more likely explanation is that artifact rejection aggressively reduced low-frequency energy of both the EP and noise. It is well established that spectral information of the ABR is composed of components from three spectral bands: 0–350 Hz, 350–700 Hz, and 700–1200 Hz (Kevanishvili & Aphonchenko, 1979; Yokoyama, 1989). Wave V is represented by a frequency region consisting mainly of low-frequency energy below 350 Hz (Suzuki, Sakabe, & Mujasbita, 1982; Yokoyama, Aoyagi, Suzuki, Kiren, & Koike, 1994). Myogenic noise resulting from physiological activity is also composed of low-frequency energy but was not a contaminating factor in this study because the analyses were limited to 7.5 ms, well below the temporal onset of any myogenic activity.

If artifact rejection destructively minimizes the EP because of its aggressive reduction in low-frequency energy, then it is possible that the phase relationships between the EP and noise interact in a detrimental way. For example, if low-frequency noise is in phase (i.e., positive) with positive ongoing wave V, it could result in an exaggerated increase in positive amplitude. This increase in amplitude would cause the sweep to be rejected if artifact rejection criterion was stringent. If, however, noise is out of phase (i.e., negative) with positive ongoing wave V, the overall sweep amplitudes would be reduced and accepted even for a strict rejection criterion. For active participants, the final average would contain an abnormally high percentage of sweeps when wave V was out of phase with the noise. This biased result would lead to an averaged ABR that had smaller than expected EPs (Stecker, 2002). Indeed, late component EPs have been reported to be smaller than normal in sleeping participants when large alpha rhythm activity interacts with the EP of interest (Rodionov & Sohmer, 2004).

We believe our results have practical clinical implications. It is generally assumed that strict rejection levels ensure SNR improvement, thus improving ABR waveform morphology. Don and Elberling (1994) found that as they lowered rejection levels from ±10 μV to ±2.5 μV, the noise was significantly reduced. However, more sweeps were needed to obtain the ABR, and the overall "quality" of the ABR waveform deteriorated. They further concluded that the SNR did not improve with systematic reductions in rejection levels, suggesting that averaging noisy sweeps may be more beneficial than averaging a subset of less noisy sweeps.

The above results were also evident in the current study for artifact rejection. Several thousand sweeps were rejected before noise RMS values were equal to Bayesian weighting, suggesting that stringent rejection levels can be time consuming and often ineffective in clinical settings. Rejected sweeps can be so numerous that a given clinical value of accepted sweeps (e.g., 2,048) may never be obtained. In fact, an example of this observation was seen in the current set of data. During the active ABR condition, the mean rejected sweeps were 4,234 (*SD* = 2,220). This clearly demonstrates that in order to achieve results comparable to Bayesian weighting, an excessive number of sweeps, sometimes thousands more, needed to be averaged. This ultimately increased test time, thereby decreasing time efficiency of the ABR. Furthermore, averaging less noisy sweeps did not guarantee better waveform morphology as seen by a reduction in wave V amplitude.

In summary, strict artifact rejection levels can lead to difficulty in ABR interpretation and could ultimately result in diagnostic errors. Our results suggest an inherent underestimation of wave V amplitude in noisy conditions when artifact rejection criteria are strictly set. Such detrimental effects contribute to reduced amplitudes and poor waveform morphology because the EP is sacrificed at the expense of the noise reduction. Upon examination of ABR morphology with AR_{10}, it was evident that waveforms appeared qualitatively flat. As mentioned above, a possible explanation that supports our observation is that some aspect of artifact rejection biases strongly against low-frequency energy, and caution should be taken when implementing stringent criteria. Future research should be directed toward evaluating whether strict rejection levels reduce important low-frequency spectral information of the ABR.

Acknowledgments

We would like to thank Drs. John Hawks, Richard Klich, Susan Motts, Kiran Nataraj, Craig Newman, and Jeff Wenstrup for their assistance on an earlier version of this article.

References

American National Standards Institute. (1996). *Specifications for audiometers* (ANSI S3.6-1996). New York: Author.

Bezerianos, A., Laskaris, N., Fotopoulos, S., & Papathahasopoulos, P. (1995). Data dependent weighted averages for recording of evoked potential signals. *Electroencephalography and Clinical Neurophysiology, 96,* 468–471.

Boston, J. R., & Ainslie, P. J. (1980). Effects of analog and digital filtering on brain stem auditory evoked potentials. *Electroencephalography and Clinical Neurophysiology, 48,* 361–364.

Boston, J. R., & Møller, A. R. (1984). Brainstem auditory evoked potentials. *Critical Reviews in Biomedical Engineering, 13,* 97–123.

Carhart, R., & Jerger, J. F. (1959). Preferred method for clinical determination of pure tone thresholds. *Journal of Speech and Hearing Disorders, 24,* 330–345.

Chandrasekhar, S. S., Brackmann, D. E., & Devgan, K. K. (1995). Utility of auditory brainstem response audiometry in diagnosis of acoustic neuromas. *American Journal of Otology, 16,* 63–67.

Coats, A. C., & Martin, J. L. (1978). Human auditory nerve action potentials and brainstem evoked responses-latency-intensity functions in detection of cochlear and retrocochlear pathology. *Archives in Otolaryngology, 104,* 709–717.

Davila, C. E., & Mobin, M. S. (1992). Weighted averaging of evoked potentials. *IEEE Translations in Biomedical Engineering, 39,* 338–345.

Don, M., & Elberling, C. (1994). Evaluating residual background noise in human auditory brainstem responses. *Journal of the Acoustical Society of America, 96,* 2746–2757.

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

Don, M., Masuda, A., Nelson, R., & Brackmann, D. (1997). Successful detection of small acoustic tumors using the stacked derived-band auditory brainstem response amplitude. *American Journal of Otology, 18,* 608–627.

Durieux-Smith, A., Edwards, C., Picton, T. W., & MacMurray, B. (1985). Auditory brainstem responses to clicks in neonates. *Journal of Otolaryngology, 4*(Suppl. 14), 12–18.

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

Elberling, C., & Wahlgreen, D. (1985). Estimation of auditory brainstem response, ABR, by means of Bayesian inference. *Scandinavian Audiology, 14,* 89–96.

Elton, M., Scherg, M., & Von Cramon, D. (1984). Effects on high-pass filter frequency and slope on BAEP amplitude, latency, and waveform. *Electroencephalography and Clinical Neurophysiology, 57,* 490–494.

Gasser, T., Mocks, J., & Verleger, R. (1983). A method to deal with trial-to-trial variability of evoked potentials. *Electroencephalography and Clinical Neurophysiology, 55,* 717–723.

Gerull, G., Graffunder, A., & Wernicke, M. (1996). Averaging evoked potentials with an improved weighting algorithm. *Scandinavian Audiology, 25,* 21–27.

Hall, J. W. (1992). *Handbook of auditory evoked responses*. Needham Heights, MA: Allyn & Bacon.

Hoke, M., Ross, B., Wickesberg, R., & Lütkenhöner, B. (1984). Weighted averaging theory and application to electric response audiometry. *Electroencephalography and Clinical Neurophysiology, 57,* 484–489.

Jacobson, J. T. (1994). *Principles and applications in auditory evoked potentials*. Needham Heights, MA: Allyn & Bacon.

John, M. S., Dimitrijevic, A., & Picton, T. W. (2001). Weighted averaging of steady state responses. *Clinical Neurophysiology, 112,* 555–562.

Kavanagh, K. T., & Franks, R. (1989). Analog and digital filtering of the brain stem auditory evoked response. *Annals of Otology-Rhino-Laryngology, 98,* 508–514.

Keselman, H. J., Keselman, J. C., & Shaffer, J. P. (1991). Multiple pairwise comparisons procedures under heterogeneous variances. *The Journal of Experimental Education, 63,* 343–362.

Kevanishvili, Z., & Aphonchenko, V. (1979). Frequency composition of brain-stem auditory evoked potentials. *Scandinavian Audiology, 8,* 51–62.
Kotlarz, J. P., Eby, T. L., & Borton, T. E. (1992). Analysis of the efficiency of retrocochlear screenings. *Laryngoscope, 102,* 1108–1112.
Lütkenhöner, B., Hoke, M., & Pantev, C. (1985). Possibilities and limitations of weighted averaging. *Biological Cybernetics, 52,* 409–416.
Marangos, N., Maier, W., Merz, R., & Laszig, R. (2001). Brainstem response in cerebellopontine angle tumors. *Otology and Neurology, 22,* 95–99.
Marsh, R. R. (1988). Digital filtering of auditory evoked potentials. *Ear and Hearing, 3,* 101–107.
Osterhammel, P. (1981). The unsolved problems in analog filtering on the auditory brainstem responses. *Scandinavian Audiology Supplement, 13,* 69–74.
Pantev, C., & Khvoles, R. (1984). Comparison of the efficiency of various criteria for artifact rejection in the recording of auditory brainstem response (ABR). *Scandinavian Audiology, 13,* 103–108.
Rodionov, V., & Sohmer, H. (2004). The contribution of the time locking of EEG waves to the generation of the auditory P300. *Journal of Basic and Clinical Physiology and Pharmacology, 15,* 71–105.
Sato, T. (1996). Type I and Type II error in multiple comparisons. *The Journal of Psychology, 130,* 293–302.
Scherg, M., & Von Cramon, D. (1984). Topographical analysis of auditory evoked potentials: Derivation of components. In R. H. Nodar & C. Barder (Eds.), *Evoked potentials II: The Second International Evoked Potentials Symposium* (pp. 73–101). Boston: Butterworth.
Schimmel, H. (1967). The (+/-) reference: Accuracy of estimated mean components in average response studies. *Science, 157,* 92.
Schwartz, D. M., & Schwartz, J. A. (1991). Auditory evoked potentials in clinical pediatrics. In W. F. Rintelmann (Ed.), *Hearing assessment* (2nd ed., pp. 429–476). Boston: Allyn & Bacon.
Sparacino, G., Milani, S., Arslan, E., & Cobelli, C. (2002). A Bayesian approach to estimate evoked potentials. *Computer Methods and Programs Biomedicine, 68,* 233–248.
Stecker, M. M. (2002). The effects of automatic rejection on evoked potential recording. *Computer in Biology and Medicine, 32,* 247–259.
Stuart, A., & Yang, E. Y. (1994). Effect of high-pass filtering on the neonatal auditory brainstem responses to air- and bone-conducted clicks. *Journal of Speech and Hearing Research, 37,* 475–479.
Suzuki, T., Sakabe, N., & Mujasbita, Y. (1982). Power spectral analysis of auditory brain stem responses to pure-tone stimulation. *Scandinavian Audiology, 11,* 25–36.
Turetsky, B. I., Raz, J., & Fein, G. (1988). Noise and signal power and their effects on evoke potential estimation. *Electroencephalography and Clinical Neurophysiology, 71,* 310–318.
Wicke, J. D., Gogg, W. R., Wallace, J. D., & Allison, T. (1978). On-line statistical detection of averaged evoked potentials: Application to evoked response audiometry (ERA). *Electroencephalography and Clinical Neurophysiology, 44,* 328–343.
Wilson, D. F., Hodgson, R. S., & Gustafson, M. F. (1992). The sensitivity of auditory brainstem response testing in small acoustic neuromas. *Laryngoscope, 102,* 961–964.
Wong, P. K. H., & Bickford, R. G. (1980). Brain stem auditory evoked potentials: The use of noise estimate. *Electroencephalography and Clinical Neurophysiology, 50,* 25–34.
Yokoyama, J. (1989). A study on three components of auditory evoked brainstem responses. *Journal of Otolaryngology Japan, 29,* 235–239.
Yokoyama, J., Aoyagi, M., Suzuki, T., Kiren, T., & Koike, Y. (1994). Three frequency component waveforms of auditory evoked brainstem response in spinocerebellar degeneration. *Acta Otolaryngology Supplement, 511,* 52–55.

Received November 15, 2005
Revision received June 1, 2006
Accepted October 10, 2006
DOI: 10.1044/1059-0889(2006/019)

Contact author: Jason Tait Sanchez, Department of Neurobiology, E108, Northeastern Ohio Universities College of Medicine, 4209 State Route 44, P.O. Box 95, Rootstown, OH 44272. E-mail: jtsanche@kent.edu.

Appendix (p. 1 of 2)

Signal Averaging Procedure

The following is a mathematical descriptive of the traditional signal averaging technique used in the current study. The terminology is as follows:

S = signal picked up from electrodes
EP = evoked potential
BN = background noise
N = number of sweeps
SNR = signal-to-noise ratio

Assuming that the EP is mixed with BN, the S was the sum of the EP and BN:

$$\mathrm{S} = \mathrm{EP} + \mathrm{BN}. \tag{1}$$

Sampled data were obtained and analyzed for the 7.5 ms of a 15-ms period (i.e., one sweep). Data obtained during the sweep were stored in the computer memory. The entire procedure was repeated *N* times. After *N* repetitions, the computers memory summed the sweeps and divided the data by the number of sweeps taken to form an averaged trace. Averaging assumes the EP to be deterministic (i.e., invariant in time and amplitude) during the sweep. Therefore, the EP is considered part of the averaged S. BN, however, is different from sweep to sweep, and after the completion of the averaging procedure, the noise amount is an averaged value of the BN. Thus, the final formula in the averaging procedure was expressed as

$$\overline{\mathrm{S}} = \mathrm{EP} + \overline{\mathrm{BN}}. \tag{2}$$

Since BN is independent of acoustic stimulus, noise contribution from sweep to sweep will tend to decrease, and therefore the averaged BN will be smaller than the BN. How much smaller depends on the number of *N* sweeps taken. Generally, the amount of noise reduction is equal to the square root of *N*:

$$\overline{\mathrm{BN}} = \mathrm{BN}/\sqrt{\mathrm{N}}. \tag{3}$$

Bayesian Weighting Procedure

Despite the use of noise reduction techniques such as filtering and artifact rejection, sweeps taken from different noise levels introduce large amounts of uncertainty in the recording of the ABR. Aware of the above problem, Hoke et al. (1984) developed a mathematical model for the EP and BN and modified the traditional signal averaging technique according to a weighted average. The equation is as follows:

$$\overline{\mathrm{S}} = \mathrm{EP} + k(\overline{\mathrm{BN}}). \tag{4}$$

The equation states that the averaged S equals the EP and the averaged BN multiplied by a factor *k,* which is a multiplication factor that best describes the BN when it is considered nonstationary (i.e., the mean and variance change over time) of the stationary zero mean BN. Under these assumptions, the weighted averaging technique theoretically estimates the true EP with a *linear least mean square error.* Another way to approach the above assumptions is based on a statistical theory known as Bayesian inference. The following mathematical explanation is adopted from Elberling and Wahlgreen (1985) and Elberling and Don (1984). Bayesian inference requires that individual sweeps used in the traditional signal averaging technique be weighted proportionally to their individual precision. According to Equation 4, the nonstationarity of the averaged BN is modeled by the multiplication factor *k,* so that sweeps are taken from a Gaussian distribution with changing variance. As information carrier of the EP, the precision of the individual sweep is inversely proportional to the magnitude of the variance. It is, however, difficult to estimate the variance based on a single sweep. Therefore, the variance of the BN is estimated based on the *signal point estimate of the BN* used by Elberling and Don (1984). The equation is as follows:

$$\mathrm{V_i} = \sum^{256} \frac{(\mathrm{V_i} - \overline{\mathrm{V}})^2}{256}, \tag{5}$$

where $\mathrm{V_i}$ is the variance of the BN based on the single point estimated of 256 sweeps (one block). The amount of BN in one sweep is obtained at a single time point within the sweep. For the current study, the time point was the 288th digitized point, or the 6 ms time point, or as follows:

$$\begin{array}{l}\text{48,000 Hz sampling rate equals 48 digitized points per millisecond (pps):}\\ \text{48 pps} \times \text{6 ms} = \text{288th digitized pps.}\end{array} \tag{6}$$

Each value of the single point estimate ($\mathrm{V_1}$) was summed and divided by 256 (one block) to obtain the mean variance of the block ($\overline{\mathrm{V}}$). The mean variance ($\overline{\mathrm{V}}$) was subtracted from ($\mathrm{V_1}$), squared, and divided by 256 to obtain the overall variance for one block of 256 sweeps ($\mathrm{V_i}$). Once $\mathrm{V_i}$ was obtained, the Bayesian estimate of the EP was as follows: After the first block (256 sweeps), we have

$$\widehat{\mathrm{EP}}_1 = (\mathrm{S_1}/\mathrm{V_1}) \times 1/\mathrm{C_1}, \quad \mathrm{C_1} = 1/\mathrm{V_1}, \tag{7}$$

where $\widehat{\mathrm{EP}}_1$ denotes the Bayesian estimate of the EP after the first block. $\mathrm{S_1}$ indicates the waveform of the first block, and $\mathrm{V_1}$ indicates the estimated variance of the first block. After the second block we have

$$\widehat{\mathrm{EP}}_2 = (\mathrm{S_1}/\mathrm{V_1} + \mathrm{S_2}/\mathrm{V_2}) \times 1/\mathrm{C_2}, \quad \mathrm{C_2} = 1/\mathrm{V_1} + 1/\mathrm{V_2}. \tag{8}$$

Appendix (p. 2 of 2)

Signal Averaging Procedure

After the *n*th block, we have

$$\widehat{EP}_n = (S_1/V_1 + S_2/V_2 \ldots + S_n/V_n) \times 1/C_n, \qquad C_n = 1/V_1 + 1/V_2 \ldots + 1/V_n \text{ or, } C_n = 1/V_n. \tag{9}$$

Equation 9 describes how Bayesian inference was used to produce an EP estimate. After division with the corresponding variances, the blocks were added together and the sum was finally multiplied by a factor $1/C_n$, which was obtained by combining all the variances. Therefore, the individual *ith* block was weighted with

$$i^{th} = n/(V_i \times C_n). \tag{10}$$

As a result, Equation 9 can be rewritten as

$$\widehat{EP}_n = 1/n \times (S_1/V_1 + S_2/V_2 \ldots + S_n/V_n) \times {}^{n}/C_n.$$

This enables comparison to be made with the traditional signal averaging technique:

$$\widehat{EP}_n = 1/n(S_1 + S_2 + \ldots S_n).$$