Evaluating residual background noise in human auditory brain-stem responses

Manuel Don
Electrophysiology Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Claus Elberling
Oticon A/S, "Eriksholm," 243 Kongevejen, DK-3070 Snekkersten, Denmark

(Received 26 May 1993; accepted for publication 29 June 1994)

The nature of the residual background noise in ABR averages was empirically examined in normal hearing objects. The residual noise in the average was estimated with use of the technique described by Elberling and Don [Scand. Audiol. **13**, 187–197 (1984)]. Low-level click stimuli were presented in 2-dB steps spanning the range from 30 to 48 dB p-p.e. SPL. For each stimulus level, 10 000 sweeps were acquired and stored for analysis. Shortcomings of the use of artifact rejection and standard averaging are demonstrated. It is further demonstrated how application of the Bayesian estimation technique of Elberling and Wahlgreen [Scand. Audiol. **14**, 89–96 (1985)] to form weighted averages can help minimize these shortcomings. Finally, the effects of smaller sweep block sizes on the Bayesian technique's ability to control the destructive effects of nonstationary noise are analyzed. Minimizing the destructive effects increases the value of statistical techniques used to detect objectively or to control the quality of ABR recordings. In all, these techniques in combination improve not only the accuracy of test interpretation but also the efficiency of clinical test time, which is becoming important for the control of medical costs.

PACS numbers: 43.64.Ri

INTRODUCTION

A. The importance and impact of background noise in research and clinical applications

In the last two decades, the use of auditory brain-stem responses (ABRs) for assessing peripheral auditory function has proliferated. ABRs have proven valuable in the audiological and neurological armamentarium. However, for most applications, a major drawback of ABRs is their low amplitudes relative to the physiological background noise which requires the use of time-consuming signal extraction techniques, such as averaging responses in the time domain. The poor signal-to-noise ratio (SNR) has been the major problem in identification of near-threshold responses and in reliable measurement of latency and amplitude of the components of the ABR for audiological, otoneurological, or neurological diagnosis. Thus the uncertainties, inaccuracies, and failures that have occurred both in research and clinical applications of the ABRs may be attributed, to a large degree, to the variable influence of the residual background noise in the averages on response measures.

B. Origin of background noise

Typically, the physiological background noise is composed of electrical activity of both neural and muscular origins. When a subject moves even slightly, large muscle potentials are often clearly observed in the recordings so that data from such an episode must be rejected. Even without overt movements of a subject, large amplitudes of background noise are frequently observed. When the background activity is large, many sweeps, often a prohibitive number, are required to reduce the noise in the average to achieve acceptable recordings. Testing may eventually be abandoned or the data are regarded as suspect. To minimize the noise levels, adult subjects are instructed to relax, close their eyes, and sleep, if possible, through the sessions. Young children often are sedated or induced to nap. Infants are tested during natural sleep, usually induced after feeding. Thus, empirically, researchers and clinicians have come to recognize that the major source of problems with ABRs is physiological background noise. They have developed guidelines for testing that focus on conditions that minimize the physiological background noise.

C. Factors affecting the SNR

The recording of evoked potentials (EPs) applies a series of different techniques to improve the signal-to-noise ratio, which is the ratio between the EP and the background noise (BN) from which the EP is extracted. The following numbered sections review briefly how the SNR can be improved by enhancing the level of the EP or reducing the level of the background noise or both. Some of the improvement techniques are "technical" and some are related to the preparation of the subject.

1. Averaging

The most widely used method is the classical technique of ensemble averaging in the time domain performed by digital signal processing in a computer. The application of this method is necessary for recording small EPs from surface electrodes. The technique computes the time average of a series of poststimulus time epochs (a number of N sweeps). If the background noise is stationary, the averaging method reduces the level of the BN in the final averaged waveform by the square root of the number of sweeps.

2. Filtering

The averaging technique is most often used in combination with frequency filtering. The purpose of filtering is to reduce or remove those components of the background noise that are present in frequency regions where the EP has no or very little energy. Because the frequency spectrum of the EP changes with stimulus parameters (e.g., level), filter settings should be selected carefully. Filters can be implemented either as analog or digital or as a combination. Analog filters are "on-line" filters and they always introduce phase delays resulting in latency shifts, amplitude changes, and waveform distortion of the EP. Digital filters are implemented in the software of the computer and are often used for "off-line" manipulations of the recorded EPs. They can be designed to have zero-phase shift whereby some of the above-mentioned shortcomings of the analog filters can be avoided. Besides improving the SNR, digital filters are often used for "smoothing" the EP waveform. The effect of both analog and digital filtering on the ABR has been described extensively in the literature (e.g., see Boston and Ainslie, 1980; Osterhammel, 1981; Marsh, 1988). Filtering is effective in improving the "global" SNR by removing noise components in frequency regions with little or no EP energy. However, because the EP and the background noise usually have overlapping frequency spectra, filtering offers only very limited improvement of the SNR, since the filter acts on both the EP and the BN and leaves the SNR unchanged. Aware of this problem, several investigators have explored various adaptive filtering techniques, e.g., the "time varying" filtering technique described by de Weerd (1981). Although their effect is sometimes significant, these techniques do not improve the detection of the EP because of the spectral nature of the ABRs and noise (Wahlgreen, 1983).

3. Artifact rejection

The averaging technique is also often combined with artifact rejection. Artifact rejection improves the SNR by excluding from the averaging process sweeps with signal amplitudes exceeding a certain preset rejection level. This effectively removes sweeps that, for instance, are spuriously affected by muscular activity from the subject. If the background noise for some reason is larger than the rejection level over a long time, no sweeps are entering the averaging process, thus the artifact rejection technique will interrupt the test. It is impossible to know the optimal setting of the rejection level before a test run because the subject's state of relaxation often changes during the test. As demonstrated in this paper, this uncertainty limits the effectiveness of the artifact rejection technique in improving the SNR.

4. Special techniques: Weighted averaging

"Weighted" averages, as developed by Elberling and Wahlgreen (1985), use a Bayesian estimation technique to reduce the destructive effects of noise variation on the ABR by weighting the average towards those blocks of sweeps with low background noise. The averaged background noise for a block of sweeps in this technique was estimated according to the procedures of Elberling and Don (1984). The present paper will further demonstrate the advantage of this technique over artifact rejection. Additionally, Hoke *et al.* (1984) and Lütkenhöner *et al.* (1985) have developed a similar method for weighting that is also based on estimates of the noise.

5. Stimulus parameters

Instead of reducing the noise to improve the SNR, the signal or EP can be increased. The EP can be enhanced by changing the stimulus parameters. Obviously, the EP can be increased simply by increasing the stimulus level. Since the EP amplitude depends on other parameters as well, they also may be chosen to enhance EPs. For example, it is well known that the stimulus type (tone burst, click, etc.) affects the EP amplitude. Although the click normally produces larger EP amplitudes than a tone burst, the click is less frequency specific and therefore could be an inappropriate stimulus. It is also well known that the EP adapts to the stimulus repetition rate and, consequently, a lower repetition rate increases EP amplitude, thereby improving the SNR. However, the lower the repetition rate, the longer is the test time for a given number of sweeps in the average. For a given test time, the gain in increased EP amplitude by use of slower repetition rates may be partially offset by higher residual noise because fewer sweeps are averaged. Therefore, the repetition or stimulus presentation rate is a parameter that must be considered to optimize a given test time.

6. Electrode placement

The magnitude of the EP depends, in part, on the placement of the two active electrodes normally used for a differential recording setup. Choice of specific electrode locations can enhance the amplitude of the EP or its specific components. Several studies have reported the influence of electrode placement on the ABR amplitude (van Olphen *et al.*, 1978; Terkildsen and Osterhammel, 1981; Parker, 1981; Starr and Squires, 1982).

7. Relaxation and sedation

The subject's state of relaxation is significant for the SNR since it controls the amount of background noise generated by the subject. To minimize this noise, the subject should comfortably recline on an inclined chair or a couch, with neck support and, if necessary, a blanket to keep warm. Children are normally given a sedative to sleep, and newborns are tested during spontaneous sleep induced by feeding.

D. Aim of the present study

This paper presents an empirical quantitative study of the background noise in ABR recordings and evaluates the effectiveness of artifact rejection and weighted averaging in minimizing the residual noise in averages. Although weighted averaging techniques have been proposed for some time, few studies utilize such techniques. This may be the result of a lack of quantitative comparisons of the benefits of a weighting scheme relative to a typical artifact rejection scheme. An important aim of this present study is to provide such comparisons with a focus on test efficiency. Furthermore, this study also explores empirically the effects of attempts to optimize weighting and rejection level schemes. In all, we hope to demonstrate the effectiveness of weighting schemes based on estimates of the physiological background noise that can improve the testing efficiency as well as the quality of the recording and, ultimately, the interpretation of ABRs.

I. METHODS

A. Subjects

Eight subjects were recruited from the staff of the House Ear Institute and House Ear Clinic. All subjects were in good general health and reported normal neurological status. Otoscopic examinations were performed to identify existing conditions that would have precluded audiometric and ABR testing. Subjects had normal hearing as defined by pure-tone thresholds at or less than 10 dB (ANSI, 1969) for frequencies between 500 and 4000 Hz and less than 15 dB for 6000 and 8000 Hz. Hearing thresholds were identified in 2-dB steps according to a modified Hughson and Westlake (1944) procedure.

B. Stimuli

Rarefaction click stimuli were produced by applying 100-μs rectangular voltage pulses to a TDH-49 earphone with an MX-41 ear cushion. Clicks were presented at regular intervals of 22 ms (approximately 45 clicks/s) and at ten levels separated by 2 dB from 30 to 48 dB peak-to-peak equivalent sound pressure level (p-p.e. SPL) with 1-kHz tone as the reference. The transduced clicks were measured and calibrated (fulfilling requirements of the American N. B. S. 9A, ANSI S3.6-1969 and IEC R303) with use of a B&K (Brüel & Kjaer (4152 artificial ear, a 6-cc coupler (DB0909), and a B&K 2209 sound level meter. Perceptual detection thresholds were determined for 1-s bursts of clicks presented through the same earphone and at the same interstimulus interval used in recording the ABRs. This perceptual threshold was defined as the 79% point on the psychometric detection function obtained in a modified block up–down procedure (Weatherill and Levitt, 1965). For the group, the average psychoacoustic threshold was 33 dB p-p.e. SPL. Thus the highest level used was about 15 dB above normal threshold. Such low levels of stimulation do not affect the background noise levels as regression analyses indicated that the estimated noise levels were independent of the stimulus level. The background noise levels for any given run depend more on the relaxation and comfort levels of the subject at the time.

C. ABR Recordings

Subjects were placed in a reclining chair in a sound-treated double-walled sound room. ABRs were obtained by recording differentially between electrodes applied to the vertex (Cz) and the ipsilateral mastoid (M1 or M2). The contralateral mastoid was used as ground. This scalp activity was bandpass filtered between 0.1 and 3 kHz with 12 dB/octave slopes and amplified such that the clipping levels of the analog-to-digital converter (ADC) corresponded to ± 10 μV at the input. The activity was sampled at a rate of 16.7 kHz for 15 ms after stimulus onset. Thus each sweep was an array composed of 256 digitized points. For each of the ten stimulus conditions, 10 000 individual sweeps were stored.

D. Data processing and analyses

The residual noise in an average was estimated according to the variance approach presented by Elberling and Don (1984). This previous work has provided detailed analyses and justification of the assumptions for this approach of estimating the background physiological noise. In that earlier paper, we considered that "... the poststimulus time epoch, $S_{(t)}$, as consisting of the evoked potential, $\mathrm{EP}_{(t)}$, and background noise, $\mathrm{BN}_{(t)}$ where the evoked potential is viewed as a deterministic signal and the background noise as a stationary, ergodic random process." The assumed Gaussian nature of background noise was tested in ten subjects and results indicate that the noise did not deviate significantly from the Gaussian assumption. Furthermore, using well-defined noises (white and pink noise), we found that the rms values obtained from a spectrum analyzer and from the single point calculation of 256 sampled values were nearly identical.

The data for this study consisted of 80 runs, ten stimulus levels for each of the eight subjects. Each run consisted of 10 000 sweeps and each individual sweep was stored. For each run, the stored individual sweeps were reprocessed to form two sets of averages: normal and weighted averages. Normal averages were formed by the straightforward process of summing the individual sweeps and dividing by the number of sweeps summed. For each of the ten stimulus conditions, seven different rejection levels were applied to the 10 000 sweeps to investigate the effect of artifact rejection levels on the normal average. These seven rejection levels were: ± 10, 8.75, 7.5, 6.25, 5.0, 3.75, and 2.5 μV. If any digitized value between 1 and 11 ms after stimulus onset exceeded the rejection level, the sweep was rejected.

Weighted averages using Bayesian estimation principles (Elberling and Wahlgreen, 1985) were formed by weighting blocks of sweeps inversely proportional to the amount of background noise estimated for that block. The mathematical expression taken from Elberling and Wahlgreen (1985) is seen in the following:

$$\widehat{\mathrm{EP}}_n = \frac{1}{n}\left(\frac{S_1}{V_1} + \frac{S_2}{V_2} + \cdots \frac{S_n}{V_n}\right)\frac{n}{C_n}, \tag{1}$$

where $\widehat{\mathrm{EP}}_n$ is the Bayesian estimate of the evoked potential after the nth block, S_i the waveform of the ith block, V_i the corresponding variance of the background noise, and C_n the sum of the reciprocal of the block variances. Thus, when the background noise (V_i) is large, that block of sweeps gets proportionally less weight in the final average. Four sets of weighted averages were formed based on the number of sweeps in a block. First, a weighted average was computed for blocks of 256 sweeps. The background noise was estimated for that block by computing the sweep-to-sweep variance of a single time point in the sweep as described by

Elberling and Don (1984). We used the time point of 5.76 ms after stimulus onset (i.e., the 96th digitized point in the sweep array). Thus for a block of 256 sweeps, 256 values were used in computing the variance and estimating the background noise. Correspondingly, the second, third, and fourth weighted averages were formed using block sizes of 128, 64, and 32 sweeps, respectively. The background noise estimates for these blocks of sweeps were computed by using, respectively, 2, 4, and 8 time points evenly spaced through the sweep. Thus 256 values were always used in estimating the background noise for the block of sweeps. These additional weighted averages based on different block sizes were used to optimize the weighting approach (see Sec. II).

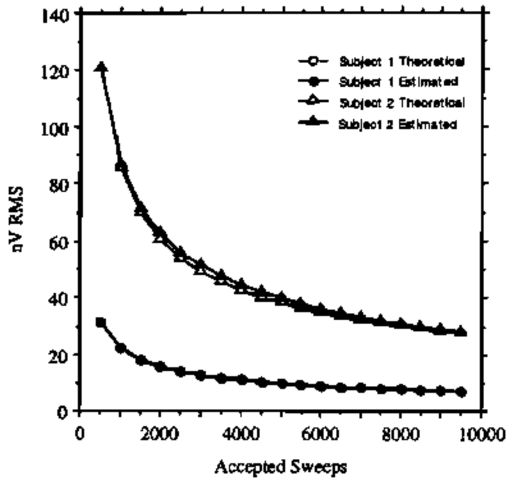

FIG. 1. Comparison of plots of the estimated residual noise and its reduction with the number of sweeps averaged and the theoretical values based on $\sqrt{N}$ between a quiet subject (subject 1) and a noisy subject (subject 2). The correspondence between the theoretical and actual average curves is related more to the high degree of stationarity rather than the low level of the noise. Even though the background noise of subject 2 was stationary, as evidenced by the similarity of the actual average and theoretical values, it took nearly 7000 sweeps to achieve the same residual noise achieved by averaging only 512 sweeps of subject 1. Furthermore, the residual noise of subject 2 after nearly 10 000 sweeps did not reduce to the level achieved by subject 1 after only 1000 sweeps. This illustrates the importance of overall noise levels.

II. RESULTS

A. Traditional averaging

Figure 1 plots for subject 1 both the estimated (filled circles) and the theoretical (open circles) residual noise for an individual run. The theoretical curve is referenced to the noise estimate after the first 512 sweeps and then reduced according to the number of additional sweeps averaged ($\sqrt{N}$). For this subject, both the estimates of the actual noise and the theoretical curve are essentially identical and cannot be distinguished. The close correspondence is due to the stationary background noise throughout the run. As evidenced by the level of the residual background noise after averaging only 512 sweeps, this subject was very quiet—in fact, sleeping.

Data from subject 2 are also plotted in Fig. 1. Subject 2 demonstrated a much higher level of background noise than subject 1. Because the background noise was also rather stationary in subject 2, the residual noise is reduced by averaging as predicted by the theoretical $\sqrt{N}$ (open and filled triangles). Although the noise was reduced by averaging as well as can be expected, the residual noise of subject 2 after nearly 10 000 sweeps did not reduce to the level achieved by subject 1 after only 1000 sweeps.

Background noise levels of subjects often do not remain stationary over the test time, as shown in Figs. 2 and 3. Figure 2 shows for subject 3 identical estimated (filled circles) and theoretical (open squares) values for the first 1500 sweeps. From 1500 to 3000 sweeps, this subject appears to be noisier as the estimates of the noise are slightly greater than the theoretical. Then the averaged residual noise rises significantly and abruptly, indicating that the subject became very noisy in the preceding block of trials. Analysis of the next block of sweeps indicates the subject became almost as quiet as before; otherwise, the observed subsequent decreases in the averaged residual noise could not have been achieved. However, even with continued averaging up to 8000 sweeps, the subject does not reach the same low noise level in the average achieved at 3000.

Figure 3 shows a similar case except that this subject (subject 4) remains noisy after the sudden, large increase in noise. In fact, the growth of the residual noise indicates that the subject's noise level is maintained at a relatively high level or is increasing over time, like a subject who awakens from sleep and whose noise level is high and/or increases. In this case, further averaging results in increasing residual noise and decreasing signal-to-noise ratio as the percentage of noisy sweeps included in the average increases. For both cases in Figs. 2 and 3, termination of the averaging before the increase in noise would have been better, but it is difficult to predict these episodic changes in background noise.

B. Bayes estimation approach

A possible solution to reducing the effect of episodic noise on the average is to form a weighted average based on the amount of noise in a block of sweeps. Such a technique was developed by Elberling and Wahlgreen (1985). The method, based on a statistical approach called Bayesian inference, forms an average that weights blocks of sweeps (e.g., 256 sweeps) inversely to the level of noise activity during the recording of that block. The averaged background noise for each block of sweeps is estimated according to the procedures of Elberling and Don (1984) and is simply computed as the sweep-to-sweep variance of a single sample point in the sweep. In Fig. 2, the application of the Bayes estimation technique is also shown on the same set of data of subject 3 discussed above. With the Bayes estimation, the averaged residual noise level does not increase as it does with normal averaging. Instead, reduction in the residual noise continues close to the theoretical value for stationary noise, whereas normal averaging never achieves the noise level before the large episodic noise burst. Similarly, in Fig. 3, the addition of Bayes approach to the same data for subject 4 is also shown. With normal averaging, the noise continues to increase; with the Bayes approach, the residual noise is maintained at the lowest level achieved before the

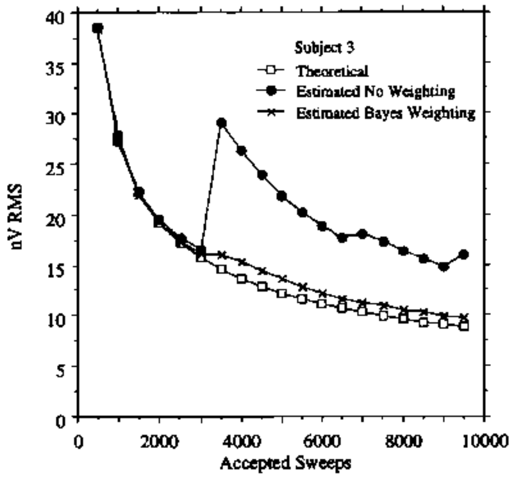

FIG. 2. A subject with a sudden increase in noise and then settling down (filled circles). Final residual noise at 9500 sweeps barely achieves level at 3000 sweeps when subject became noisy. However, applying the Bayesian weighting approach (×'s) reduces the detrimental effect of the episodic noise and noise reduction continues close to the theoretical values (open squares) based on $\sqrt{N}$.

large increase in the background noise. This prevents the average waveform from becoming more contaminated by noise and thereby prevents decreasing the signal-to-noise ratio as averaging continues.

C. Artifact rejection level approach

Perhaps the most common approach to reducing noise when averaging is to reject sweeps with high levels of noise. The criterion for rejection is typically any sweep in which one or more sample points exceed a defined voltage value in a specified time of the sweep. For example, if the range of the ADC is ±5 V, which, after accounting for amplifier gain, corresponds to ±10 μV from the scalp, a rejection level of ±3 V means that any sweep whose activity in a given time epoch of the sweep exceeds ±6 μV at the scalp is to be

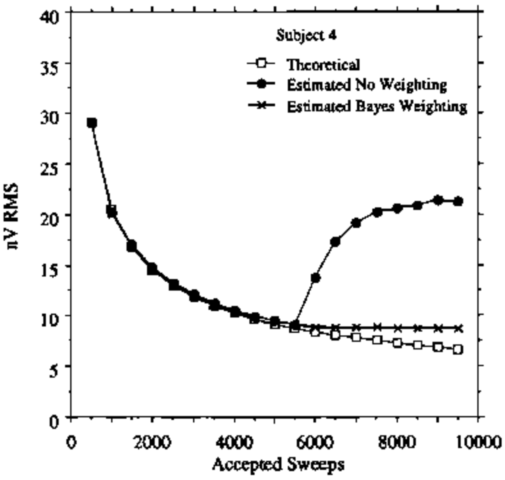

FIG. 3. Estimated residual noise as a function of number of sweeps. As the subject became noisy at around 6000 sweeps, estimated residual averaged noise level (filled circles) continues to increase. However, the Bayesian weighting approach (×'s) prevents an increase in the residual noise and maintains the noise level achieved before the increase.

rejected. The disadvantage of rejection levels lies in determining the appropriate value for the given level of background noise. Too stringent a rejection level can mean many rejected sweeps, thereby increasing the number of sweeps processed and the testing time. Figure 4(a) demonstrates the effect of rejection level on the averaged background noise. The same set of 10 000 sweeps is processed according to different rejection levels and the estimated residual noise in the average is computed. The data are plotted as a function of accepted sweeps. Clearly, the lower the rejection level, the lower the averaged background noise for a given number of accepted sweeps obtainable at each of the rejection levels. This result is expected since the averages with lower rejection levels are formed from an equal number of less noisy sweeps. However, as the rejection level becomes more stringent, the maximum number of accepted sweeps becomes smaller. For example, at the low rejection level of ±3.75 μV (filled circles), only about 2500 sweeps were accepted (arrow at "a") after processing 10 000 sweeps; i.e., only one in four sweeps was accepted. Yet the residual noise level was higher than for the less stringent ±5.00 μV (×'s) rejection level (arrow at "b"), where slightly more than 5500 sweeps were accepted after processing the same 10 000 sweeps. The ±5.00 μV level resulted in lower averaged noise than the ±3.75 μV because more sweeps were accepted. This illustrates the problem with rejection levels: The more stringent the rejection level, the greater the number of total sweeps required to form an average composed of a given number of accepted sweeps. At the most stringent rejection level (±2.5 μV) examined, no data are plotted because fewer than 256 sweeps were accepted after processing the 10 000 sweeps at this rejection level.

To determine the true value of artifact rejection, a plot of the averaged background noise as a function of the number of sweeps processed, i.e., total number of stimulus presentations, is shown for the same set of data in Fig. 4(b). Since the number of stimulus presentations determines clinical test time, the required total number of sweeps processed (i.e., presented), rather than the number of accepted sweeps to reduce the averaged noise to a given level, should be the criterion for judging test efficiency. When the data are plotted in this manner, there appears to be no advantage to using artifact rejection levels for this run. The lowest noise for a given number of sweeps or stimulus presentations is obtained with use of the highest (i.e., least stringent) rejection level (±10 μV input, corresponding approximately to the clipping level of the ADC). For any given number of sweeps processed, the stringent rejection levels of ±5.0 and ±3.75 μV yielded greater residual noise in the average than less stringent rejection levels. Thus, for this set of data, averaging more, yet noisier, sweeps appears more advantageous than averaging a subset of less noisy sweeps.

D. Artifact rejection levels versus Bayesian weighting approach

The results in Fig. 4(b) beg the question of whether artifact rejection level helps at all in terms of processed sweeps. To evaluate the value of artifact rejection level in normal averaging, the following analysis was performed.

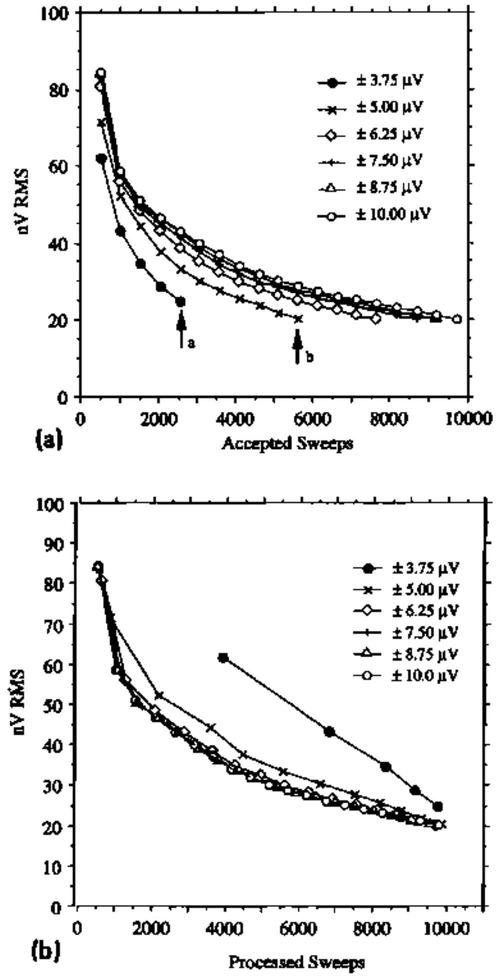

FIG. 4. (a) The effect of varying rejection level on the estimated residual noise as a function of **accepted** sweeps is a reduction of the noise with increased stringency of the rejection level. However, after all sweeps are processed, a less stringent rejection level (± 5 μV, arrow b) achieves lower residual noise than the more stringent rejection level (± 3.75 μV, arrow a). (b) Data are a function of **processed** instead of **accepted** sweeps. More stringent rejection levels are less efficient in reducing the residual noise.

The computed residual noise level in the average after processing 10 000 sweeps was used as a reference value. Only sweeps that exceeded the clipping level of the ADC (approximately ± 10 μV) were rejected. A rejection level was defined as having a significant effect if the residual noise level was reduced by at least 5% relative to the reference value. The 5% noise reduction corresponds to an approximately 10% improvement of the F_{sp} value which was the criterion used by Elberling and Wahlgreen (1985) for claiming a significant benefit of their Bayesian weighting scheme. Figure 5(a) and (b) summarizes the results of applying various levels of artifact rejection. For the 80 runs evaluated in this study, the scatter plot in Fig. 5(a) shows that many runs had a significant reduction of the residual noise when some level of artifact rejection smaller than the comparison level of ± 10 μV was used. However, several runs showed an increase in residual noise when a smaller rejection level was applied. Also shown in Fig. 5(a) are the results of the Bayes approach. In this approach, it is unnecessary to determine an appropriate rejection level because all sweeps except those that exceed the clipping level of the ADC are used. Note that in contrast to results with artifact rejection levels, *no* runs resulted in greater residual noise with the application of Bayesian weighting to the same data set.

Figure 5(b), from the same data set as Fig. 5(a), plots the percentage of runs in which the residual noise was (1) reduced significantly (by 5% or more), (2) unchanged (less than $\pm 5\%$ change), or (3) increased significantly (by 5% or more). This figure shows how the distribution of these results varied with rejection level. As the rejection level was decreased, the percentage of runs in which the noise was reduced by 5% or greater generally increased, with the maximum percentage (28%) occurring for the ± 3.75 μV rejection level. However, the percentage of runs showing an increase in residual noise also grew dramatically as more stringent rejection levels (e.g., ± 3.25 and ± 2.5 μV) were applied. On the basis of these results, it is obvious that the rejection level of ± 2.5 μV is too stringent: The number of runs showing sufficient noise reduction becomes smaller while the number of runs showing noise increase becomes larger. In some cases, after processing 10 000 sweeps, this rejection level did not permit the estimate of the background noise for even one block of trials because the required number of accepted sweeps (256) was not attained. Overall, the higher residual noise in the final average results from fewer (although quieter) sweeps accepted for averaging.

In contrast, as noted before, no runs with the Bayes approach showed an increase of any magnitude. All runs (56%) classified as no significant change were either equal (0% change) or were slightly reduced (i.e., greater than 0% but less than 5% reduction). Second, the percentage of runs (44%) that resulted in significant noise reduction [filled circles in Fig. 5(b)] was greater for the Bayesian weighting than for any of the rejection levels used. Figure 5(c) shows the quantitative distribution of the percentage noise reduction for the Bayes results in Fig. 5(a): 35 of the 80 runs (44%) showed noise reduction of 5% or more, with a median percent noise reduction of about 16%. A 16% noise reduction corresponds to an improvement of the F_{sp} value of about 35% and, therefore, is equivalent to the improvement achieved by normal averaging of 35% more sweeps. for comparison, Fig. 5(d) shows the distribution of applying the "best" (in terms of percentage of runs resulting in criterion noise reduction) artifact rejection level of ± 3.75 μV. Fewer runs were reduced and the median value of the amount of noise reduction was about 12%. Notably, several runs showed residual noise increases.

A major problem with using artifact rejection levels is determining *a priori* the rejection level that will reduce the noise without compromising the averaging efficiency. Figure 6 demonstrates the difficulty in selecting an appropriate rejection level for three subjects showing different rejection level functions. In this figure, the effects of rejection levels on the percent change in the residual noise are shown. Subject 2A (open squares) shows the desired effect in which the more stringent the rejection level, the greater the reduction (negative % change) in the estimated residual noise. However, as noted previously, increased stringency of the rejection level will not consistently reduce the residual noise for a

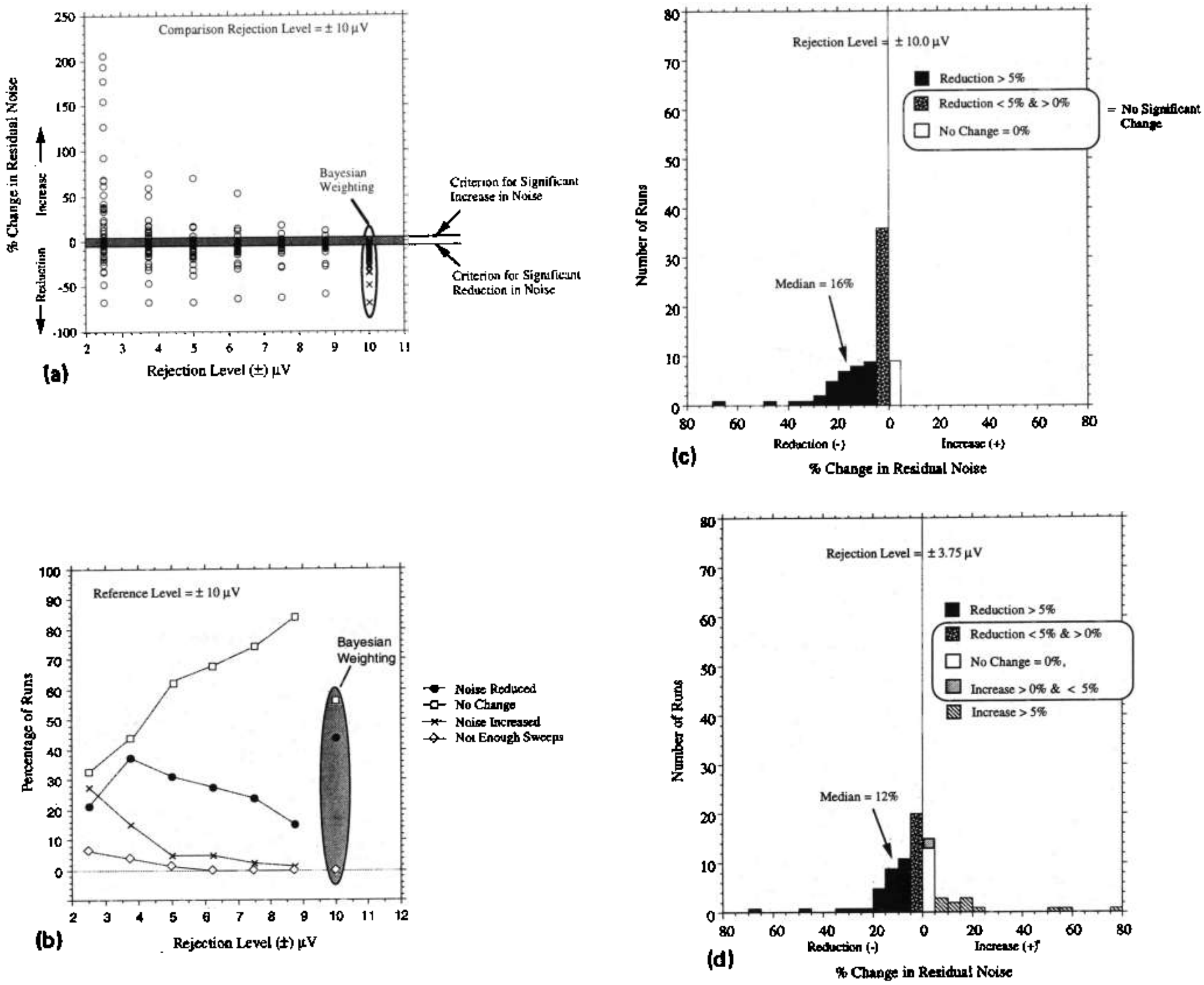

FIG. 5. (a) Scatter plot of the change in residual noise for all rejection levels relative to rejection at the clipping level of the ADC, which corresponds to ± 10 μV at the input. Noise reductions are noted by negative % changes and noise increases, by positive values. Criterion for significant noise increase or reduction is a change of 5% or more. At all rejection levels there are some runs showing an increase in the residual noise after processing 10 000 sweeps. The more stringent the rejection level applied, the greater the number of runs showing increase in the estimated residual noise. However, the results of Bayesian weighting on the same data set show no runs with increased residual noise. (b) Summary of the effect of rejection level on the residual noise for all runs. Criterion for significant noise reduction or increase is a change of 5%. The percentage of runs with increased residual noise becomes larger with more stringent rejection levels. Applying Bayesian weighting reveals no runs with increase noise and 44% of the runs with significant noise reduction. This percentage is greater than for any rejection level applied. (c) Distribution of the percentage residual noise reduction for Bayesian weighting results in (a). (d) Distribution of the percentage residual noise reduction for "best" artifact rejection level (± 3.75 μV) results in (a).

given number of processed sweeps. Subject 3A (filled circles) shows all three possible outcomes (no significant change, significant reduction, and significant increase) in the residual noise depending on the rejection level. Subject 7A ($\times$'s) shows no benefit from increased rejection level but rather a monotonic increase in the residual noise as the rejection level is increased in stringency. These variations are not subject specific as these three patterns can occur on three different runs from the same subject. Also shown in Fig. 6 are the results applying the Bayesian weighting scheme to the same data for the three subjects. The reduction of the residual noise is greater than achieved by any of the rejection levels. Thus, using the weighting approach, one can achieve better noise reduction without concern for determining the appropriate rejection level. One simply sets the rejection level at ± 10 μV, which after amplification corresponds to the clipping level of the ADCs.

E. Stationarity of background noise

Figure 5(b) showed that nearly 55% of the runs were unaffected by the Bayesian weighting approach. It is likely that runs with no change in the residual noise with Bayesian weighting exhibit stationary noise irrespective of the actual noise level. In contrast, runs with little effect of the application of stringent rejection levels must have both low and stationary background noise.

To determine if the advantage of the Bayes approach relates to the stationarity of the background noise in terms of episodic variations, a simple measure to quantify these varia-

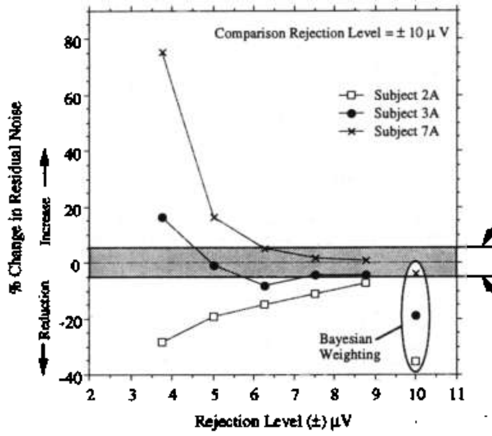

FIG. 6. Variable effects of artifact rejection level on the residual noise in the ABR average for three different subjects. In comparison, the Bayes weighting approach (circumscribed symbols) resulted in lower residual noise than the best rejection level for all three subjects.

tions was developed. The residual background noise for each block of 256 accepted sweeps is computed as described in Sec. I. A given run of 10 000 sweeps with little or no clipped sweeps yields approximately 39 blocks of 256 sweeps. Thus there are 39 noise estimates, one for each block. The mean for these 39 block noise estimates (MBN_b) is computed and divided by the standard deviation (SDBN_b) of these 39 block noise estimates. This ratio defines our simple measure of the degree of block-to-block stationarity. Thus

$$\text{degree of stationarity} = \frac{\mathrm{MBN}_b}{\mathrm{SDBN}_b}.$$

Statistically, this ratio is the inverse of the coefficient of variation. In essence, high stationarity throughout the run means that the computed residual background noise from one block of sweeps to the next will be similar and the standard deviation will be relatively small. Thus the ratio of the mean block background noise and the standard deviation for the blocks will be large. For runs that have poor stationarity, i.e., the computed residual noise from one block to the next varies greatly, the standard deviation for the blocks will be relatively large and the ratio will be small. Figure 7(a) plots the relationship between the degree of noise stationarity as defined above and the amount of change in residual noise achieved by Bayesian weighting. Runs with a degree of stationarity value below 6 show significant noise reduction and a clear systematic relationship between the amount of noise reduction and the degree of stationarity. The less the degree of stationarity, the greater the effect of Bayesian weighting on reducing the residual noise. Above 6 or so, the stationarity is sufficiently high to minimize the effect of the Bayesian weighting.

Figure 7(b) shows the relationship between noise reduction and mean noise amplitude of the blocks. As expected, there is no systematic relationship since the Bayesian weighting technique should be sensitive to stationarity rather than amplitude of the noise. Thus the Bayesian weighting technique is effective in reducing the residual noise when the

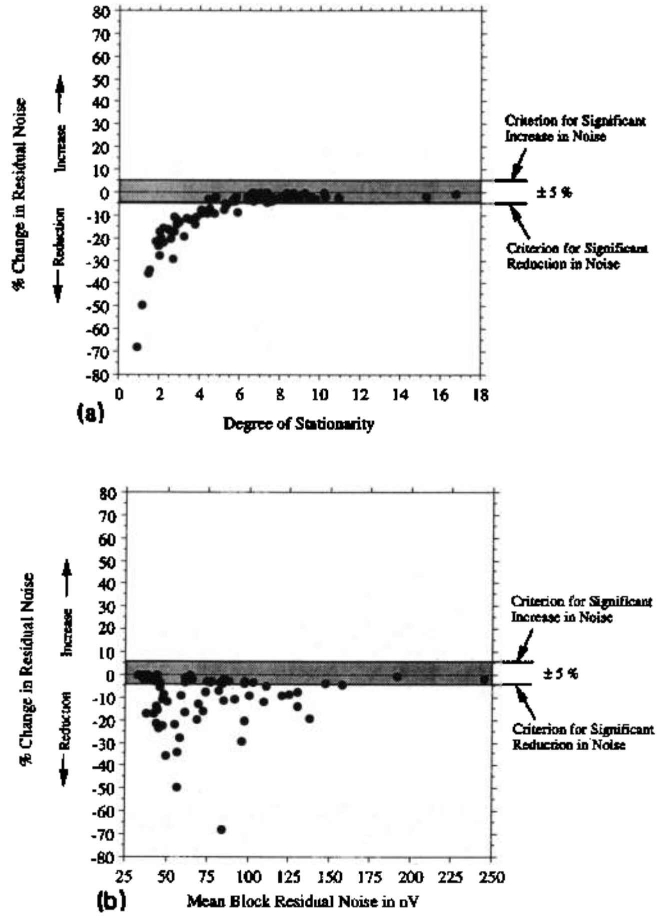

FIG. 7. (a) The relationship between the amount of noise reduction achieved by the Bayesian weighting approach and the degree of stationarity of the background noise throughout a run. The lower the degree of stationarity, the greater the percentage of noise reduction obtained with weighting. As expected, the weighting approach only minimally improves runs with high degrees of stationarity. (b) No apparent relationship exists between the amount of noise reduction achieved by the Bayesian weighting approach and the overall level of the background noise in a run as expressed by the mean block residual noise.

background noise is nonstationary across blocks of sweeps independent of actual level.

Figure 8(a) plots the relationship between the degree of noise stationarity and residual noise change when a rejection level of ±3.75 μV is applied. This rejection level was chosen for comparison because its application produced a significant reduction in the residual noise in more runs than any other rejection level. Several observations are noted. First, a fairly systematic relationship exists between the amount of noise reduction and degree of stationarity when the value is 3 or less; above 3, results are unpredictable. Second, it appears that most of the runs that showed increased residual noise were in the stationarity degree range of 3–8. Finally, little change in the residual noise is seen for degrees of stationarity greater than 8. Thus other factors in addition to stationarity affect the results of applying artifact rejection. Figure 8(b) shows the relationship of the noise reduction with artifact rejection at ±3.75 μV and mean noise amplitude of the blocks. Unlike the Bayes weighting, there appears to be a significant positive correlation ($r=0.44, p<0.0001$) between the mean block noise and the percent change in residual

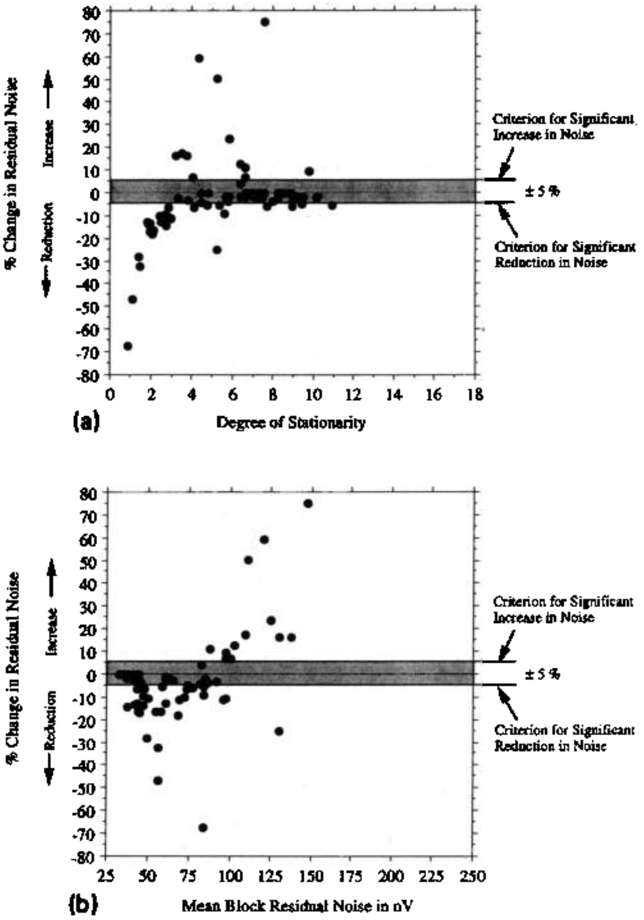

FIG. 8. (a) The relationship between the amount of noise reduction achieved by the artifact rejection approach and the estimated background noise throughout a run. Although a definite relation appears for low degrees of stationarity, this relationship is unpredictable at slightly higher degrees of stationarity and often results in a substantial increase in residual noise. The Bayesian approach [Fig. 7(a)] does not produce these unpredictable results. (b) Unlike the results for the Bayesian approach shown in Fig. 7(b), there seems to be a relationship between the amount of noise reduction with rejection level and the overall noise in the run as expressed by the mean block residual noise. The estimated residual noise seems to be reduced for low noise levels and increased for higher noise levels.

noise. The most likely reason for this effect is that with increased noise levels, more sweeps are rejected and cannot contribute to the averaging process to reduce the noise. The less stringent the rejection level, the poorer this relationship. However, the effect of artifact rejection on noise reduction is not related in a simple linear way to noise amplitude.

F. Optimization of Bayesian weighting approach

The Bayes approach uses estimates of the background noise computed as the variance of a single sample point from swept to sweep in a block of 256 sweeps. Since a minimum of 200 independent samples allow a good estimate of the background noise (Elberling and Don, 1984), the 256 values from a block of sweeps are more than adequate. The drawback is that the whole block of 256 sweeps is then weighted as a unit. With large variations in a block of sweeps, the weighting factor computed may be inappropriate for many sweeps within the block. The ideal weighting scheme would weight each sweep, but this is prohibitive because an accurate estimate of the background noise cannot be achieved with a single sweep. However, at least 8 degrees of freedom usually exist in the average waveform (Elberling and Don, 1984). Thus, eight independent, equally spaced sample points may be taken per sweep to estimate the noise. This would reduce the number of sweeps in a block required to estimate the noise and permit better weighting. For example, if eight sample points can be taken from each sweep and assumed to be independent, a block size of 32 sweeps would provide 256 independent sample points from which to estimate the background noise. An estimate of the noise would still be accurate but the weighting factor would apply to only 32 instead of 256 sweeps. The advantage of weighting smaller blocks of sweeps would be better control of the noise variation because smaller block sizes are closer approximations to weighting the single sweep. Theoretically, the residual noise in the final average formed by weighting small blocks of sweeps should be less than that formed by weighting the larger 256 sweep blocks. To test whether weighting smaller blocks of sweeps will reduce the average noise significantly in practice, the same set of data was processed with use of the Bayes approach but with estimates of the background noise from two, four, and eight sample points per sweep and weighting of blocks of 128, 64, and 32 sweeps per block, respectively. The estimated residual noise in the average from the different block sizes was compared to the normal single sample point per sweep and 256 sweep block size. In all conditions, 256 sample points were used to estimate the background noise in the Bayes approach weighting, and all averages used not only the same number but also the same sweeps except that they were weighted differently. The effect of block size optimization of the Bayesian weighting was evaluated after 4096 accepted sweeps.

Figure 9(a) shows in histogram form the results of varying the block size on the residual noise. For each of the three different block sizes, the percentages of runs in which the residual noise changed by various amounts relative to that obtained with the standard 256 sweeps per block are plotted. All three block sizes appear to have a significant portion of runs resulting in increased residual noise. The smallest block size (32 sweeps/block) does appear to have the most runs that resulted in some noise reduction. Figure 9(b) plots the mean data for each of the three block sizes for all runs, for only those runs in which Bayesian weighting reduced the noise by criterion level (i.e., 5% or greater reduction), and for runs that did not achieve criterion reduction. Several aspects to these data are notable. First, the smaller the block size, the greater the reduction of the residual noise. To avoid the issue of distribution, nonparametric statistics were used to assess whether the percent change in residual noise as a function of block size differs significantly from 0 (i.e., no change). Results of application of the one sample sign test, shown in Table I, indicate that the reduction caused by block sizes 32 and 64 was significant beyond the 0.0001 and 0.0111 levels, respectively. Specifically, for block sizes of 32 sweeps, 49 (62%) of the 78 runs showed a reduction in the residual noise whereas only 12 (15%) showed an increase and 17 (22%) showed no change. However, the amount of noise reduction is very small, as the mean is at best not more

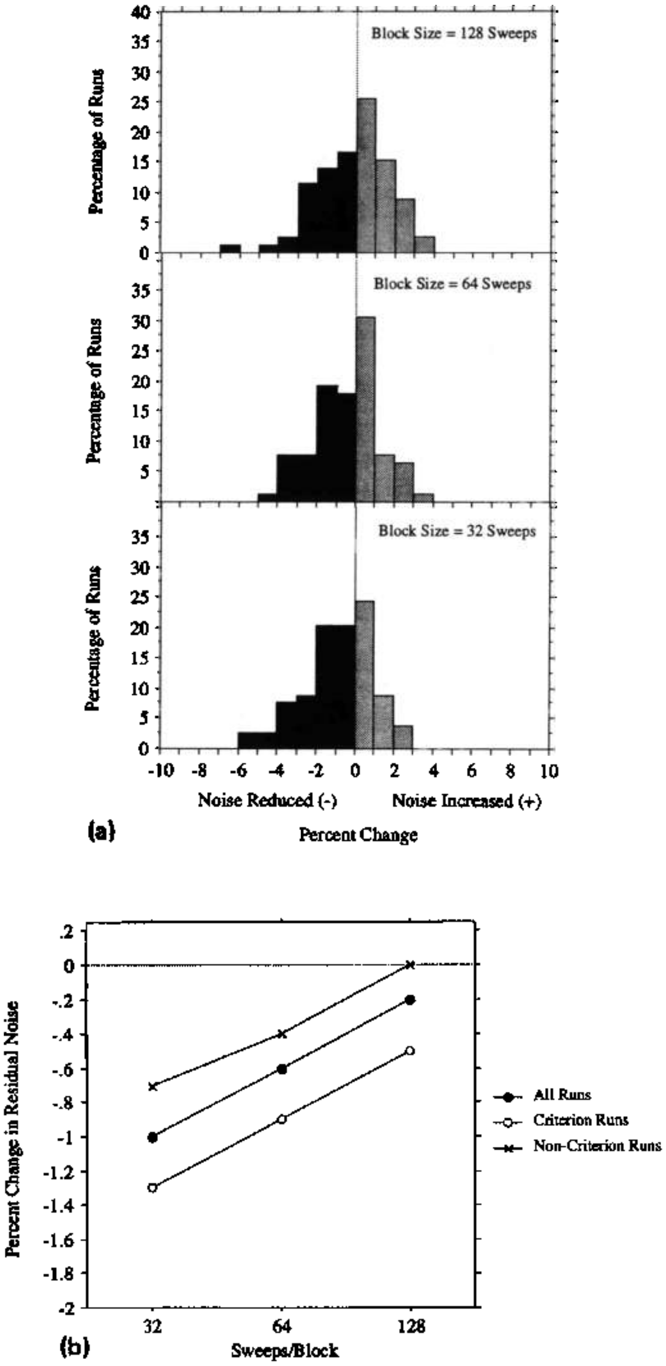

FIG. 9. (a) Histograms of the effect of block size on changes in the residual noise. The reference is the block size of 256 sweeps. The percentage of runs with reduction seems to increase with smaller block sizes. (b) Mean effect of block size on residual noise in final average. The effect for all runs and the runs separated on the basis of whether criterion noise reduction was achieved by the Bayesian weighting approach are compared.

than 1.4% and the maximum reduction for a run was about 5%–6%. The block size of 128 did not cause a statistically significant reduction ($p=0.389$).

Second, both Fig. 9(a) and (b) also suggests that the reduction differs for different block sizes. The Wilcoxon signed rank test (nonparametric paired t test), shown in Table II, indicates that there is a significant although very small difference ($p<0.05$) among all block sizes. Third, Fig. 9(b) also suggests that the amount of reduction is greater for runs in which the Bayesian weighting was more effective. To test whether a correlation exists between the amount of reduction with the Bayesian weighting technique and the amount of additional reduction with use of block size optimization, Spearman rank correlation was performed. Results in Table III show that for both the 32 and 64 sweep block sizes, the improved reduction appears to be significantly ($p<0.05$) correlated with the amount of reduction from the Bayesian weighting application.

TABLE I. One-sample sign test of the effect of block size (BS) on noise reduction. Hypothesized value=0.

	BS=128	BS=64	BS=32
No. obvs>hyp. value	29	21	12
No. obvs<hyp. value	37	42	49
No. obvs=hyp. value	12	15	17
P value	0.389	0.011	<0.0001

In summary, reduced block sizes apparently improves the Bayesian weighting scheme, although the mean amount of improvement is very small. The amount of noise reduction is greater for smaller block sizes and is correlated to the amount of reduction achieved by Bayesian weighting.

III. DISCUSSION

A fundamental precept of examining averaged electrical activity for ABRs is that the averaged waveform is composed minimally of averaged background noise. Whether an ABR also exists in the averaged waveform depends on the degree of stimulation. Whether the ABR can be detected in the averaged waveform depends also on the degree of stimulation, ratio of the ABR to residual background noise levels, and the criteria used for judging the presence of an ABR. Variations in threshold measures as well as other parametric measures of component peaks in the ABR can be understood more fully by evaluating the nature of the residual averaged background noise. A simple variance calculation can be used to estimate the residual background noise level in an ABR average, and such estimates can be used to quantify estimates of signal-to-noise ratios for objective, statistically based detection of ABRs (Elberling and Don, 1984, 1987a,b; Don *et al.*, 1984). Variations in physiological background noise can often significantly affect both visual and/or statistical detection of ABRs. Destructive effects cannot always be efficiently minimized by simple averaging. This evaluation of the physiological background noise touches on several issues to improve our interpretation of ABRs and our efficiency in administering these tests to patients.

Obviously, the most important issue is the overall noise level (Fig. 1). Since the frequency composition of the background noise and ABR overlap, filtering schemes are limited in the amount of SNR improvement, thus requiring reliance

TABLE II. Wilcoxon signed rank test of the difference between block sizes (BS).

BS 128 vs BS 32	P value<0.0001
BS 128 vs BS 64	P value<0.014
BS 64 vs BS 32	P value<0.0001

TABLE III. Spearman rank correlation between amount of noise reduction related to block size (BS) and noise reduction related to Bayesian weighting.

BS 128	ρ value=0.174	P value=0.1276
BS 64	ρ value=0.275	P value=0.0156
BS 32	ρ value=0.287	P value=0.0119

on response averaging. In theory, the noise is reduced as the square root of the number of sweeps averaged. From the perspective of testing time, the required averaging increases with the square of the noise level. For example, if the input noise is a factor of 3 larger, theoretically nine times as many sweeps are required to reduce the residual noise to comparable levels. In a practical situation, such as testing in a clinic, an estimation of the noise level would greatly aid the course of testing. A high and constant noise level may require an inordinate amount of averaging to reduce the residual noise in the final average to permit either visual or statistical detection of an evoked response with reasonable confidence. Thus a quantitative estimate of the noise level allows the assessment of the attainability of a response with a reasonable signal-to-noise ratio.

A second issue is the frequent departure of the noise from stationarity; that is, movement or changes in the subject's arousal state will alter the level of the background noise. Episodic increase in the background noise can cause a dramatic increase in the residual noise level of the average. Such an increase would require thousands of additional sweeps to reduce the residual noise to the level just before the episodic noise occurred (Fig. 2). Another possibility is that the subject awakens or becomes more restless and noisier such that the SNR of the average continues to decrease with time (Fig. 3). In such a case, the ABR waveform may appear to deteriorate with continued averaging. The typical approach to minimizing the influence of episodic increase in noise has been the use of artifact rejection schemes. It is important to remember that the efficiency of a technique to reduce the residual noise should be evaluated in terms of the amount of test time required for the noise reduction, that is, the number of sweeps processed (i.e., total stimuli presented) rather than the number of sweeps accepted for the average. According to this measure of efficiency, varying the rejection level is not a totally satisfying strategy. Figure 5(a) and (b) indicates that any rejection level below 10 μV (input corresponding to the clipping level for the amplifier) will result in some runs with more residual noise than obtained with the 10 μV rejection level. Although more stringent artifact rejection levels can often reduce the residual noise, they can also increase residual noise for a given number of stimulus trials because many sweeps will be rejected and too few sweeps will be averaged for reducing the noise. In general, the number of runs with an increase in residual noise increases as the rejection level becomes more stringent. A few runs had a continuously high background noise level such that application of the rejection level that overall produced the most noise-reduced runs resulted in acceptance of less than one block of 256 sweeps out of 10 000 sweeps [Fig. 5(b)].

Moreover, the effect of rejection level on residual noise can vary widely among individuals ranging from continued reduction to continued increase with increasing stringency of the rejection level (Fig. 6). Too lenient a level may still allow sweeps with destructive noise levels to be averaged. Thus, the dilemma is to decide at what level of rejection does the exclusion of the sweep outweigh the benefit of its inclusion. Elberling and Wahlgreen (1985) demonstrated that this problem can be solved by applying Bayesian estimation principles to form an average in which blocks of sweeps are weighted inversely to the estimated average noise level in the blocks. A decision on appropriate rejection level is eliminated, as only sweeps that exceed the clipping level of the amplifier are rejected.

There are several advantages to the application of this technique. First, weighting blocks of sweeps inversely to the estimated background physiological noise level controls the influence of large episodic noise (Fig. 2) or of increasing noise (Fig. 3). Second, a comparison of this method with standard artifact rejection level showed that, in the same set of data, the Bayesian weighting technique never resulted in increased residual noise levels for any run, whereas artifact rejection did so for one or more runs at all rejection levels [Fig. 5(a)]. Third, the Bayesian weighting technique reduced significantly the residual noise in more runs than that achieved by any rejection level [Fig. 5(b)]. Fourth, the need to guess at the rejection level is eliminated as one simply sets it to the clipping level of the amplifying system. Finally, as shown by Elberling and Wahlgreen (1985), if a technique such as the F_{sp} (Elberling and Don, 1984, 1987b; Don *et al.*, 1984) utilizing a statistical criterion for ABR detection related to the SNR is used, detection criterion will be achieved in fewer sweeps with the Bayesian weighting technique than with any rejection level.

Our experience in this data set indicates that the residual background noise was reduced appreciably (5% or greater) in more than 40% of the runs. This percentage of runs is consistent, although slightly larger, with that of Elberling and Wahlgreen (1985), who reported a similar benefit in more than 30% of their runs.

Obviously, a constant noise level—either low or high—would yield little advantage to weighting blocks of sweeps since each block would essentially be weighted the same and would be equivalent to simple averaging. Thus the Bayesian weighting approach is valuable for improving runs mainly having episodes of background noise bursts. The reduction in the residual noise is greatest for runs defined as having low degree of stationarity [Fig. 7(a)]. For artifact rejection levels, the relationship is not as well defined [Fig. 8(a)]. Similar to the Bayesian weighting, the greatest noise reduction also occurred for the runs with the lowest degree of stationarity. Moderate degrees of stationarity seem to produce a preponderance of runs with higher residual noise. At high degrees of stationarity, little change in the residual noise is observed with artifact rejection. Complicating the picture with artifact rejection is the apparent sensitivity to overall noise level [Fig. 8(b)]. Such sensitivity is not observed [Fig. 7(b)], nor theoretically expected, with a weighting scheme.

The attempt to optimize the Bayes approach by weighting smaller blocks of sweeps yielded minimal improvement. The hypothesis was that smaller block sizes would yield even lower noise estimates because smaller block sizes approach the ideal of weighting each individual sweep. Single sweeps cannot be individually weighted because an accurate estimate of the noise would be difficult and time consuming. The limit of how small a block size can be used depends on the expected number of degrees of freedom in the sweep to allow use of independent samples for estimating the background noise. The smallest block size used was 32 sweeps, from which eight evenly spaced sample points per sweep were taken. This yielded 256 points for the variance calculation used in estimating the background noise. Overall, as theoretically expected, the smaller the block size, the better the noise reduction. The greatest improvement occurred with the smallest block size [Fig. 9(a) and (b)], but this block size reduced the residual noise in the average by only 1% more than the standard 256 sweep block size. Nonparametric statistics indicate that the improvement, although small, was significant for the block sizes of 64 and 32. Furthermore, the different block sizes produced significantly different results. Finally, the amount of residual noise reduction from decreasing the block size was shown to be significantly correlated with the amount of noise reduction resulting from the Bayesian weighting application to the standard block size of 256. In all, the lack of strong improvement with the reduction of sweep block size in the residual background noise with the Bayes approach suggests that the episodic noise tends to occur over the time spanning a block of 256 sweeps. Since this study used approximately 45 sweeps per second, this translates to about 5.6 s. However, the significant correlation between the amount of noise reduction with the Bayesian weighting and the amount of additional reduction with block size suggests a relationship of the nonstationarity across blocks and the degree of nonstationarity within a block of sweeps.

The frequent observance that the residual noise actually increased with a reduction in block size is to be expected, since the calculation of the estimate of the background noise for the smaller block size uses more than one point per sweep. Thus one can expect variations in the weighting factor, particularly when there is no short lasting episodic noise.

Although the advantage of weighting smaller blocks of sweeps is minimal, it may be worthwhile to implement if it does not compromise test time and if the appropriate degrees of freedom exist to permit the use of multiple sample points in a sweep for estimating the background noise.

IV. SUMMARY

Nearly all previous ABR studies have referred to the variations in the latency and amplitude parameters of response component peaks and troughs. While indeed there are likely to be actual variations in the responses themselves, we feel that a large portion of the variations from run to run within and across subjects are attributable to unaveraged residual noise. We examined empirically the nature of the residual background noise in ABR averages and demonstrated some shortcomings of standard averaging and the use of rejection levels. We applied the Bayesian estimation technique of Elberling and Wahlgreen (1985) to demonstrate quantitatively how this approach can help to minimize these shortcomings and to control the destructive effects of episodic noise. Minimizing the destructive effects of episodic noise increases the value of statistical techniques used to detect objectively or to control the quality of ABR recordings. In all, these techniques in combination improve not only the accuracy of test interpretation but also the efficiency of clinical test time, which is becoming important to the control of medical costs.

ACKNOWELDGMENTS

We thank Ann Masuda for collecting the data and helping with data processing. We also thank the reviewers for their helpful comments. This work was supported by NIH Grant No. 1R01 DC00043.

ANSI. (1969). ANSI S3.6-1969, "American National Standard Specifications for Audiometers" (ANSI, New York).

Boston, J. R., and Ainslie, P. J. (1980). "Effect of analog and digital filtering on brainstem auditory evoked potentials," Electroencephalogr. Clin. Neurophysiol. 48, 361–364.

de Weerd, J. P. (1981). "A *Posteriori* time-varying filtering of averaged evoked potentials. I. Introduction and conceptual basis," Biol. Cybernet. 41, 211–222.

Don, M., Elberling, C., and Waring, M. W. (1984). "Objective detection of averaged auditory brainstem responses," Scand. Audiol. 13, 219–228.

Elberling, C., and Don, M. (1984). "Quality estimation of averaged auditory brainstem responses," Scand. Audiol. 13, 187–197.

Elberling, C., and Don, M. (1987a). "Threshold characteristics of the human auditory brain stem response," J. Acoust. Soc. Am. 81, 115–121.

Elberling, C., and Don, M. (1987b). "Detection functions for the human auditory brainstem response," Scand. Audiol. 16, 89–92.

Elberling, C., and Wahlgreen, O. (1985). "Estimation of auditory brainstem responses, ABR, by means of Bayesian inference," Scand. Audiol. 14, 89–96.

Hoke, M., Ross, B., Wickesberg, R., and Lütkenhöner, B. (1984). "Weighted averaging—Theory and application to electric response audiometry," Electroencephalogr. Clin. Neurophysiol. 57, 484–489.

Hughson, W., and Westlake, H. (1944). "Manual for program outline for rehabilitation of aural casualties both military and civilian," Trans. Am. Acad. Opthamal. Otolaryngol. Suppl. 48, 1–15.

Lütkenhöner, B., Hoke, M., and Pantev, C. (1985). "Possibilities and limitations of weighted averaging, " Biol. Cybernet. 52, 409–416.

Marsh, R. R. (1988). "Digital filtering of auditory evoked potentials," Ear Hear. 9, 101–107.

Osterhammel, P. (1981). "The unsolved problems in analog filtering on the auditory brain stem responses, " Scand. Audiol. Suppl. 13, 69–74.

Parker, D. J. (1981). "Dependence of the auditory brainstem response on electrode location," Arch. Otolaryngol. 107, 367–371.

Starr, A., and Squires, K. (1982). "Distribution of auditory brainstem potentials over the scalp and nasopharynx in humans," Ann. N.Y. Acad. Sci. 388, 427–442.

Terkildsen, K., and Osterhammel, P. (1981). "The influence of reference electrode position on recordings of the auditory brainstem responses," Ear Hear. 2, 9–14.

van Olphen, A. F., Rodenburg, M., and Verwey, C. (1978). "Distribution of brainstem responses to acoustic stimuli over the human scalp," Audiology 17, 511–518.

Wahlgreen, O. (1983). "Time varying filtering of ABR," M.Sc. thesis, Laboratory of Acoustics, Technical University of Denmark.

Weatherill, G. B., and Levitt, H. (1965). "Sequential estimation of points on a psychometric function," Br. J. Math. Stat. Psychol. 18, 1–10.