Use of the Median Method to Enhance Detection of the Mismatch Negativity in the Responses of Individual Listeners

Lisa G. Fox*
Susan D. Dalebout*

Abstract

The median method was evaluated as an alternative way of expressing the mismatch negativity (MMN). Traditionally, signal averaging has been used to extract these event-related potentials from unwanted background noise. However, mean values are biased by unrejected artifact that skews the relatively small distribution of values on which the MMN is based. Because the median is a more valid measure of central tendency in asymmetric distributions, it may describe MMN data more accurately. Better representation of the signal in the median waveform might enhance detection of the MMN in the responses of individual listeners. Mean and median waveforms were computed from previously recorded MMN data. Visually identified MMNs were validated using area and onset latency criteria. Detectability of the MMN was not improved using median waveforms. Despite this result, a theoretical argument for use of the median is presented.

Key Words: Auditory evoked potentials, event-related potentials, median method, mismatch negativity, signal averaging

Abbreviations: ABR = auditory brainstem response; AEP = auditory evoked potential; Data Set 1 = 30 records obtained in a contrast condition and 30 records obtained in a control (no contrast) condition; Data Set 2 = 20 records obtained in a contrast condition; ERPs = event-related potentials; MMN = mismatch negativity; SNR = signal-to-noise ratio

Sumario

El método de la mediana fue evaluado como una forma alternativa de expresar la "negatividad desigual" (mismatch negativity: MMN) Tradicionalmente se ha utilizado la promediación de la señal para extraer estos potenciales relacionados con el evento (event-related potentials) del ruido de fondo no deseado. Sin embargo, los valores promedio están afectados por artefactos no rechazados que sesgan la distribución relativamente pequeña de valores en los que está basada la MMN. Dado que la mediana es una medida más válida de la tendencia central en el caso de distribuciones asimétricas, ésta puede describir en una forma más exacta la información relacionada con la MMN. Una mejor representación de la señal en la onda promedio podría incrementar la detección de la MMN en las respuestas de sujetos individuales. La media y la mediana de las ondas fue calculada a partir de información sobre la MMN previamente registrada. Fueron validados los valores de la MMN visualmente identificados utilizando criterios de área y latencia de inicio (onset latency). La detectabilidad de la MMN no mejoró por el uso de la mediana de las ondas. A pesar de estos resultados, se presenta un argumento teórico a favor del uso de la mediana.

Palabras Clave: Potenciales evocados auditivos, potenciales relacionados con el evento, método de la mediana, negatividad desigual, promediación de la señal

Abreviaturas: ABR = respuesta auditiva de tallo cerebral; AEP = potencial evocado auditivo; Data Set 1 = 30 registros obtenidos en condición de contraste y 30 registros obtenidos en una condición de control (sin contraste); Data Set 2 = 20 registros obtenidos en condición de contraste; MMN = negatividad desigual; SNR = relación señal/ruido

*Communication Disorders Program, University of Virginia, Charlottesville, Virginia
Reprint requests: Lisa G. Fox, Communication Disorders Program, University of Virginia, 2205 Fontaine Avenue, Suite 202, Charlottesville, VA 22903

Signal averaging has been used to express auditory evoked potentials (AEPs) since the late 1940s (Hall, 1992). This technique involves calculation of mean (ϵX/n) voltage across repeated trials at each sampling point in a recording epoch. Theoretically, the invariant "signal" (i.e., neurophysiologic activity time-locked to the acoustic stimulus) accumulates as the average, whereas random "noise" (e.g., ongoing, spontaneous activity unrelated to the stimulus) cancels out. That is, noise is just as likely to be positive as negative from one trial to the next and will sum to zero, provided that enough trials are averaged. By extracting the signal from unwanted background noise, signal averaging *should* enable detection of the AEP in the mean waveform.

In practice, noise is not eliminated completely from the averaged response. Nonetheless, accurate signal detection can occur as long as the signal-to-noise ratio (SNR) is sufficiently positive (i.e., the signal is greater than residual noise). The adequacy of the SNR depends on the degree to which the following conditions are met (Hyde, 1994): (1) the signal is invariant, (2) noise is random and stable over time, (3) the signal and noise are independent, (4) noise in successive trials is statistically independent, and (5) noise is normally distributed. Although AEPs will violate one or more of these conditions, degradation of the SNR is problematic only when residual noise obscures the presence of a signal. For example, the auditory brainstem response (ABR) is a relatively invariant AEP that is somewhat resistant to noise contamination due, in part, to the large number of trials on which the average is based. Because the signal is highlighted adequately in the mean waveform, the ABR can be detected effectively even in the responses of individual listeners. In contrast, detection of event-related potentials (ERPs) in the averaged responses of some listeners can be difficult.

ERPs reflect sound processing that occurs at higher levels of the auditory system (relative to the ABR). Elicitation of ERPs is contingent on the functional significance of the evoking stimuli, circumstances of stimulus presentation, and/or listener participation (Chertoff et al, 1988; Kraus and McGee, 1994). For example, the mismatch negativity (MMN) is an ERP that reflects automatic processing of stimulus differences in the absence of listener attention, whereas the P_{300} is an ERP that reflects conscious, task-oriented discrimination of contrasting stimuli. Because ERPs are sensitive to fluctuations in psychophysical state (e.g., level of arousal, attention to the stimulus, habituation), they are more variable than obligatory responses generated at lower levels of the auditory system. That is, ERPs are vulnerable to both "random" (i.e., changing from trial to trial) and "systematic" (i.e., progressing over time) forms of variation (Hyde, 1994). As a consequence, all five conditions can be violated, in which case, a poor SNR might interfere with ERP detection.

For example, violation of the first condition (i.e., the signal is invariant) is evidenced by slight trial-to-trial variation in ERP latency (i.e., jitter). The signal also can vary systematically, such as when prolonged repetition of a stimulus induces habituation of the response (Hyde, 1994; Picton et al, 2000). Both types of variation blur representation of the signal in the averaged waveform, reducing ERP amplitude and deteriorating waveform morphology. Suppression of the signal makes detection of the ERP more difficult because of the increased likelihood that the response will be masked by residual noise.

The next three conditions are violated by physiologic variability that is unrelated to the stimulus. For example, background noise (e.g., eye blinks, electromyogenic noise, alpha rhythms) is neither consistent nor random (as dictated by the second condition). Rather, the magnitude and/or frequency composition of this unwanted activity is likely to change over time. Moreover, background noise can repeat at intervals correlated with stimulus presentation and/or the ERP itself (Hall, 1992). Because this is likely to occur when the ERP signal and noise have related electrophysiologic sources and/or common sources of modulation (Hyde, 1994), the third condition (i.e., the signal and noise are independent) also is violated. The fourth condition (i.e., noise in successive trials is statistically independent) is violated when low-frequency noise persisting across trials results in a significant correlation between trials (Hyde, 1994; Sinkkonen and Tervaniemi, 2000). Violation of these conditions complicates ERP detection because relatively more noise is preserved in the mean waveform.

The small sample size associated with ERPs makes it probable that the fifth condition (i.e., noise is normally distributed) will be violated as well. ERPs such as the MMN and P_{300} are collected in an oddball paradigm in which the stimulus eliciting the response occurs far less frequently than a repeating standard stimulus (e.g., ratio = 15:85). Thus, the number of responses containing the signal is relatively small. In addition, the total number of trials collected in ERP paradigms is reduced by the slower rate of stimulus presentation required to record these responses. Because of the relatively small number of trials on which ERPs are based (as compared to the ABR, for example), the distribution of voltages representing each sampling point is likely to be asymmetric. As a result, less noise is "averaged out" of the ERP waveform because it sums (cumulatively) to some value other than zero. In addition, infrequent artifacts (i.e., unrelated electrical activity of relatively large voltage) have a large impact on ERP waveforms because these extreme values are less likely to be cancelled out in a small sample (Hall, 1992; Yabe et al, 1993; Hyde, 1994). Unfortunately, collection of additional trials to improve the SNR is precluded by practical limits on session duration and the threat of confounds such as listener fatigue and/or habituation.

Given the limitations of signal averaging, Borda and Frost (1968) proposed an alternative data reduction technique that accounted for signal variability, noise contamination, and sample size. This pioneering work was largely ignored until the last decade, when it was revisited by researchers seeking to improve P_{300} detection in the responses of individual listeners (Yabe et al, 1993; Yabe, 1997). The apparent success of these experiments is intriguing, particularly in light of the difficulty associated with identifying the MMN in the responses of individual listeners (McGee et al, 1997; Ponton et al, 1997; Dalebout and Fox, 2000, 2001; Picton et al, 2000; Tremblay et al, 2001). Dalebout and Fox (2001) recently recommended that techniques for enhancing the MMN detectability be explored; in this follow-up effort, an alternative method for expressing ERPs was investigated.

Median Method

Borda and Frost (1968) first asserted that it is more appropriate to express AEPs based on a small number of trials in terms of median (rather than mean) voltage. For a rank-ordered set of values, the median is defined as the value above and below which 50 percent of the values fall, that is, the 50th percentile (Glass and Hopkins, 1996). As demonstrated in Figure 1, the mean and median are equivalent when the distribution of values is symmetric; however, such distributions are typically achieved only with large sample sizes. Relatively smaller samples, like those associated with ERPs, are more likely to have skewed (i.e., asymmetric) distributions in which the mean value is drawn toward the tail by extreme values. Although more than 50 percent of the values fall above or below the mean in a negatively or positively skewed distribution (respectively), the median reflects the center of the distribution regardless of its shape. Similarly, the mean is more affected by the absolute number of extreme values in the distribution (i.e., degree of kurtosis) because these values inflate the cumulative sum (ϵX) contained in its numerator.

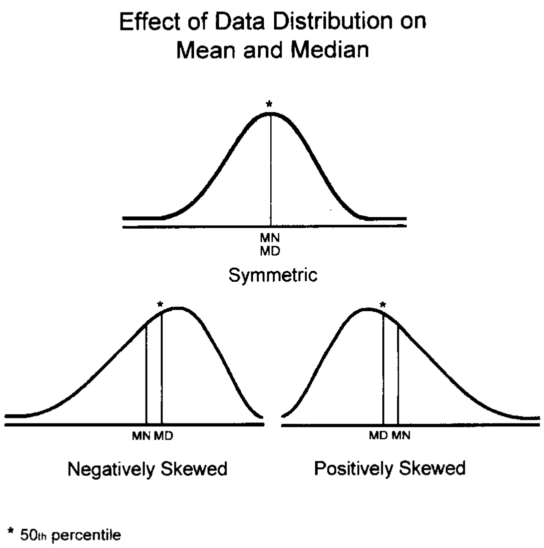

Figure 1 In a symmetric distribution, the mean (MN) and median (MD) are equivalent; both represent the 50th percentile. However, in negatively and positively skewed (i.e., asymmetric) distributions, the mean is pulled toward the tail by extreme values.

Although artifact rejection techniques commonly are used to exclude trials contaminated by unrelated neurophysiologic activity, such methods minimize (rather than eliminate) extreme values in the ERP average. For example, all trials exceeding a predetermined voltage criterion (e.g., ± 100 μV), such as those related to ocular movement, can be rejected online. However, infrequent artifacts that do not reach the specified criterion (e.g., amplitude = ± 99 μV) can contaminate the averaged waveform. Because the median is less affected by extreme values, it is a better estimator of common voltage (i.e., the signal) in a skewed distribution.

To illustrate this point, Yabe and colleagues (1993) constructed mean and median waveforms from a model using artificial signals of constant amplitude (peak amplitude = ± 20 μV;

n = 30) and white noise of varied amplitude (SD = 0 to ± 20 μV). Although application of a single rectangular pulse (amplitude = + 300 μV) was reflected in mean waveforms as a large positive deflection, median waveforms were unaffected by this infrequent artifact. Similarly, these authors demonstrated that unrejected ocular artifact contained in ERP data for a real listener inflated the mean waveform such that amplitude appeared artificially large. In contrast, the median waveform was unaffected. These findings have important implications for the clinical application of ERPs. For example, comparison of ERP amplitude to normative values for the purpose of identifying a clinical condition (e.g., auditory processing disorder) could be confounded by the influence of "outliers" (i.e., nonrepresentative members of the distribution) on mean values. That is, inclusion of extreme values in the mean computation could bias the estimate such that outcome (i.e., whether or not the ERP is identified) is inaccurate (Hyde, 1994).

The median method has also been used to address the most severe form of signal variability, missing signals. Using the same model described above, Yabe and colleagues (1993) systematically manipulated the percentage of trials containing the signal (0–100%). Comparison of mean and median waveforms demonstrated that median amplitude was consistently larger when the signal was probable (i.e., occurring in more than 50% of trials). Interestingly, the opposite result was obtained (i.e., mean amplitude was larger) when the signal was improbable (i.e., occurring in less than 50% of trials). This finding is relevant to ERPs given the possibility that a signal may not be generated in response to each stimulus (or, in the case of an oddball paradigm, each "target" stimulus) because of psychophysical instability. Assuming that this type of "error" does not occur more than 50 percent of the time, these findings suggest that ERP amplitude would be more accurately depicted by the median waveform. Moreover, the observation that median amplitude was larger for probable signals in all noise conditions prompted the authors to conclude that the median method is "more sensitive than the averaging method in finding signals concealed by background activity" (p. 407).

In another compelling demonstration, Yabe (1997) simulated trial-to-trial latency jitter using both artificial and real P_{300} data. By shifting signal onset forward and backward in time, the author evaluated the differential effect of latency variation on mean and median waveforms. Median amplitude was consistently larger than mean amplitude, indicating that latency jitter had less impact on signal representation when the median was used to reflect central tendency.

Although the median method has been shown to be a viable alternative for expressing the P_{300} in the responses of individuals (Yabe et al, 1993; Yabe, 1997), there have been no reports of its application with MMN data. This is somewhat surprising, given that the median method may be better suited for use with the MMN than with the P_{300}. That is, the probability of signal occurrence should be higher for the MMN because it is recorded in the absence of listener attention and is not subject to signal variability associated with task execution. Furthermore, because the amplitude of the MMN is comparably small (due to the absence of listener attention during recording), it is easily masked by electrophysiologic noise (Picton et al, 2000). As a result, the MMN is considerably more difficult to detect than the P_{300} in the responses of individual listeners. If the median method can be used to isolate the MMN from background noise by minimizing the contribution of infrequent artifacts and maximizing signal amplitude, MMN detectability might be enhanced. The purpose of the current investigation was to compare mean and median waveforms derived from the same data to evaluate the viability of the median method as an alternative means for expressing the MMN.

METHOD

Data

A total of 80 records from two previously collected MMN data sets were reanalyzed (Dalebout and Fox, 2000, 2001). A record constitutes a set of responses from an individual listener elicited by a particular stimulus contrast. Although stimulus and recording parameters for the data sets were identical, they were analyzed separately for two reasons. First, 30 of the 60 records forming "Data Set 1" (Dalebout and Fox, 2000) were collected in a control (i.e., no contrast) condition that allowed performance characteristics of the MMN (i.e., hit rate, false alarm rate, and d') to be evaluated. Second, the 20 records forming "Data Set 2" (Dalebout and Fox, 2001) consisted of 4 records from each of five listeners. Thus, records obtained from each listener were not independent; theoretically, outcome in 1 record should be correlated with outcome in the other 3 records obtained from the same listener.

Stimuli

Standard and deviant stimuli were taken from a nine-item synthetic speech continuum varying in place of articulation from /dɑ/ to /gɑ/. Stimulus parameters have been described in greater detail elsewhere (Dalebout and Stack, 1999; Dalebout and Fox, 2000). Briefly, the syllables differed according to the starting frequency of their first and second formant (F1, F2) transitions. The continuum was created by systematically varying the starting frequencies of F1 and F2 in equal steps from the end point of /dɑ/ (step 1) to the end point of /gɑ/ (step 9). The continuum end points (i.e., steps 1 and 9), which are maximally contrastive (i.e., constitute the greatest acoustic difference along the continuum), formed the stimulus contrast used to elicit the MMN. As expected, most listeners easily discriminated the 1 to 9 contrast in a two-alternative, forced-choice same/different task (mean percent correct = 94%). Step 9 was presented alone in the control condition of Data Set 1.

Listeners

The records represent a total of 35 listeners (mean age = 23 years): 30 from Data Set 1 (Dalebout and Fox, 2000) and 5 from Data Set 2 (Dalebout and Fox, 2001). All listeners were female college students who reported normal neurologic function and no history of learning problems. Audiologic screening ensured normal hearing sensitivity bilaterally.

Data Acquisition

Procedures for recording MMN data have been described in previous reports (Dalebout and Fox, 2000, 2001). Neuroscan software (STIM and SCAN programs) and hardware (Synamps) were used for stimulus presentation and data collection. Stimuli were presented to the right ear at approximately 72 dB SPL through an insert earphone; listeners listened to the audio portion of a self-selected videotaped film (~ 40 dBA) in free field. The interstimulus interval was 1.1 seconds, and the ratio of standard to deviant stimuli was 85:15 in the oddball sequence used for contrast conditions. Activity was recorded from electrodes at Fz and Cz; the nose served as reference and the forehead as ground. Supraorbital and infraorbital electrodes arranged in a bipolar montage around the left eye monitored eye artifact. Using a 500 points/second sampling rate and a recording epoch of 550 msec (including 100-msec prestimulus time), each response consisted of 275 data points. Responses were collected in trial blocks of 500 or 1000; each record consisted of approximately 1800 to 2000 responses after online artifact rejection ($\pm$ 100 μV). The data were analog filtered online from 0.1 to 100 Hz.

Procedures

Mean and median waveforms were computed for all records in Data Sets 1 and 2 and examined for the presence of an MMN. Onset latency and area-under-the-baseline were measured for all negative deflections identified as possible MMNs. Validation criteria were established using mean waveforms from Data Set 1 and were applied to all waveforms (mean and median) in both data sets to determine the rate of MMN identification.

Computation of Waveforms

Each trial block for a given record was sorted into two *.eeg* (Neuroscan) subsets according to response type. Responses to deviant stimuli constituted one subset and responses to standard stimuli immediately preceding deviants constituted the other subset (Schröger, 1998; Sinkkonen and Tervaniemi, 2000). Corresponding subsets were appended sequentially across trial blocks (e.g., deviant1 + deviant2...) to generate composite files of standard and deviant responses. Approximately 270 to 300 responses were included in each composite file; sample sizes for standard and deviant files were similar.

When preparing the data for export, all electrodes were "skipped" except the one chosen for analysis. A utility program (*Eeg2asc2.exe* developed by Neuroscan) was used to convert each composite file to an ASCII file. Because this utility exports *.eeg* data points into a single column (i.e., rows = points), a macro was created in Microsoft Excel to write data points across columns; that is, each row represented a single response (i.e., number of rows = number of responses). Although each response consisted of 275 data points, it was necessary to delete the final 19 data points (38 msec) because an Excel spreadsheet contains only 256 columns. Finally, data were pasted into a different Excel workbook that computed mean and median values on separate spreadsheets. Importantly, each measure of central tendency was based on the same set of responses; thus, mean and median values represented equivalent sample sizes.

Importation of mean and median data points into Neuroscan required several steps. First, the *Paste Special* command was used in Excel to replace formulas with computed (mean or median) values and to transpose the row containing these values into a single column consisting of 256 data points. Second, the spreadsheet was saved in text (tab-delimited) format. Third, the file was renamed in MS-DOS with a *.dat* extension. Finally, the ASCII file was opened in Neuroscan's *Edit* program and saved as an average file (regardless of whether it contained mean or median data).

Following importation, each standard waveform was subtracted from its corresponding deviant waveform to derive a difference waveform. Finally, all waveforms were digitally lowpass filtered at 20 Hz and linearly detrended (for an explanation see Dalebout and Fox, 2001).

Analysis of Waveforms

Waveforms from Data Set 1 were coded such that both examiners were unaware of the condition (i.e., contrast vs control) in which the data had been collected. Blind analysis minimized bias during visual identification of MMNs. Because the MMN can overlap the N1 of the auditory late response (Alho, 1995; Picton, 1995; Sinkkonen and Tervaniemi, 2000), its onset would be expected to occur sometime after P1 (i.e., the component immediately preceding N1). Therefore, a minimum criterion for MMN onset latency was established a priori; the onset of a negative deflection in the difference waveform had to exceed 60 msec (i.e., the approximate latency of P1; Tucker et al, 2001) to be considered a possible MMN.

Measurement of Area-under-the-Baseline

The latency at which the difference waveform crosses the baseline marks the intersection of overlaid standard and deviant waveforms (i.e., 0 μV difference). These latencies were identified as the onset and offset of the visually identified MMN. Waveforms were sent to the *Waveboard* in Neuroscan's *Edit* program to enable precise measurement. When a data point measuring exactly 0 μV could not be identified, the latency of the first negative data point following the onset or preceding the offset was used.

Once the onset and offset latencies were recorded, the waveforms were saved as ASCII files and exported to a spreadsheet. Data points for each waveform were checked against a key to ensure that only those with negative voltages (i.e., under the baseline) were included in the computation. When multipeaked MMNs were identified, the area for each negative deflection was computed separately; multiple areas then were summed on a spreadsheet (to guard against rounding errors).

Derivation of Validation Criteria

Once decisions regarding the presence or absence of the MMN had been made, records were decoded and categorized by stimulus condition (i.e., contrast or control). Validation criteria were established using mean waveforms from Data Set 1 (which included the requisite control condition). Procedures for deriving these criteria have been explained in greater detail elsewhere (Dalebout and Fox, 2000, 2001). Briefly, a combination of MMN area and onset latency criteria was used to validate MMN responses (McGee et al, 1997). Each criterion was defined by the value at which d' was optimized. d' is a statistic that incorporates both hit rate (i.e., percentage of presumably correct identifications of the MMN in a contrast condition) and false alarm rate (i.e., percentage of incorrect identifications of the MMN in a control condition) to indicate the accuracy of a test. As such, d' is useful for comparing test performance (i.e., hit rate and false alarm rate) as a cutoff criterion is systematically manipulated. Criteria for area and onset latency were varied separately, with area considered first and onset latency considered second. That is, only those negativities exceeding the area criterion were included in the second analysis that determined the criterion for onset latency. MMN area was given greater weight because McGee and colleagues considered it the best single measure for validating MMN responses. In the present investigation, d' was optimized when negative deflections with an area-under-the-baseline greater than 29 msec $\times$ μV and an onset latency less than 217 msec were identified as MMNs. Once established, these criteria were used to validate visually identified MMNs in mean and median waveforms from both data sets.

Table 1 Hit Rates, False Alarm Rates, and *d′* Values Associated with MMN Identification among the Responses of Individual Listeners

	Contrast Condition		*Control Condition*		
	MMNs Identified	*HT (%)*	*MMNs Identified*	*FA (%)*	*d′*
Data Set 1					
Mean	18/30	60	7/30	23	1.00
Median	14/30	47	8/30	27	.54
Data Set 2					
Mean	9/20	45			
Median	9/20	45			

MMN = mismatch negativity; HT = hit rate; FA = false alarm rate.

RESULTS

Data Set 1

Hit rates, false alarm rates, and *d′* values for Data Set 1 are shown in Table 1. The larger *d′* value for mean waveforms (1.00), as compared with median waveforms (0.54), is attributable to a higher hit rate (60%) and slightly lower false alarm rate (23%). The 14 MMNs identified in median waveforms were not a subset of the 18 MMNs identified in mean waveforms; rather, there were records in which an MMN was identified in the median but not the corresponding mean waveform and vice versa.

Data Set 2

When the same validation criteria were applied to the waveforms forming Data Set 2, an identification rate of 45 percent (9/20) was obtained for both mean and median waveforms (see Table 1). Although nine MMNs were identified using each method, they did not necessarily represent the same records. That is, there were two records for which an MMN was identified in either the mean or median waveform (but not both).

Table 2 Number of Mismatch Negativities Identified per Listener in Data Set 2

	No. of MMNs Identified	
Listener	*Mean*	*Median*
1	2	2
2	2	2
3	4	3
4	1	2
5	0	0

As shown in Table 2, the number of MMNs identified among the four records for each listener ranged from 0 to 4 for mean waveforms and 0 to 3 for median waveforms. Although three listeners had the same identification rate for mean and median waveforms, different identification rates were obtained for two listeners. One listener failed to demonstrate an MMN with either method of data expression.

Waveform Characteristics

Mean area and onset latency were calculated for MMNs validated in mean and median waveforms recorded in both contrast (i.e., hits) and control (i.e., false alarms) conditions. These values are summarized in Table 3, as are the ranges and standard deviations. No marked differences in MMN magnitude or onset latency were noted as a function of waveform type or stimulus condition. However, examination of individual waveforms showed that MMNs validated in both mean and median waveforms from the same record often differed in area and/or morphology. For example, mean area was greater than median area on occasion, whereas the reverse was true in other cases. In addition, median waveform morphology generally was poorer (i.e., appeared more "noisy") than mean waveform morphology. As a result, MMN identification was judged (by the examiners) to be more difficult in median waveforms.

DISCUSSION

The median method was investigated as an alternative to signal averaging for expressing MMN data collected from individual listeners. Use of the median did not enhance MMN detection in this study. It is likely that MMN identification was limited by a poor SNR, regard-

Table 3 Ranges, Means, and Standard Deviations for Area-under-the-Baseline and Onset Latency

	Data Set 1				*Data Set 2*	
	Hits		*False Alarms*			
	Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*
Area*						
Range	29–237	39–180	38–206	33–182	43–192	53–124
Mean	91	99	105	98	100	96
SD	52.8	40.2	57.7	48.1	47.1	27.9
Onset latency†						
Range	62–214	72–211	71–210	77–211	87–196	108–215
Mean	137	131	121	141	135	153
SD	48.8	37.7	46.3	49.6	35.3	45.3

*Area is expressed in msec × μV.
†Onset latency is expressed in msec.

less of the manner in which waveforms were expressed. That is, the signal was too small and/or the noise was too great to permit detection of the MMN in the responses of many listeners. Unfortunately, detection was too poor to make a comparison of the mean and median methods meaningful.

Although a slightly larger d' value was obtained for mean waveforms ($d' = 1.00$) than for median waveforms ($d' = 0.54$), it is possible that this difference is an artifact created by use of mean waveforms to maximize d'. That is, the accuracy of MMN detection might have been better for median waveforms had they been used to derive the criteria for MMN validation. However, any difference in outcome likely would have been negligible.

Mean and Median Deviant Waveforms Representing the Same Data

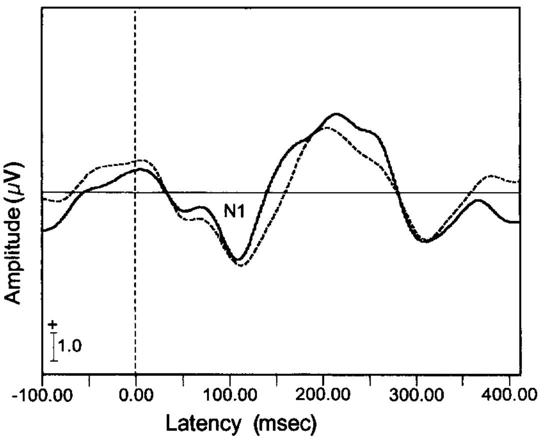

Figure 2 Deviant waveforms from one record in Data Set 1; the mean waveform (*dotted line*) and median waveform (*solid line*) are overlaid. The N1 component of the auditory late response is clearly observed in both waveforms. However, divergence of mean and median waveforms indicates asymmetry of the underlying voltage distributions.

In any case, the poor performance characteristics (i.e., hit rate, false alarm rate) obtained for mean and median waveforms would prevent use of the MMN as a diagnostic tool. For example, mean and median waveforms were associated with hit rates of 60 percent and 47 percent, respectively. Given that the MMN was not detected in a large number of responses recorded from normal listeners, its absence must be considered within the range of "normal." Picton and colleagues (2000) noted that when "the limits of normal amplitude include zero, it will not be possible to consider any MMN abnormally small or absent" (p. 132). Likewise, the false alarm rates obtained for mean (23%) and median (27%) waveforms are unacceptable for diagnostic testing. That is, identification of an MMN when no response is present (i.e., in a control condition) implies that genuine absence of an MMN, perhaps associated with neuropathology, could be missed. As a result, the utility of the MMN for individual assessment is limited at the present time (Dalebout and Fox, 2000, 2001; Kraus and Cheour, 2000; Picton et al, 2000; Tremblay et al, 2001).

Despite poor detection of the MMN in the present study, the theoretical premise supporting use of the median remains credible. The observation that mean and median waveforms are different when overlaid would indicate that the data distribution underlying each sampling point (in the region of disparity) is

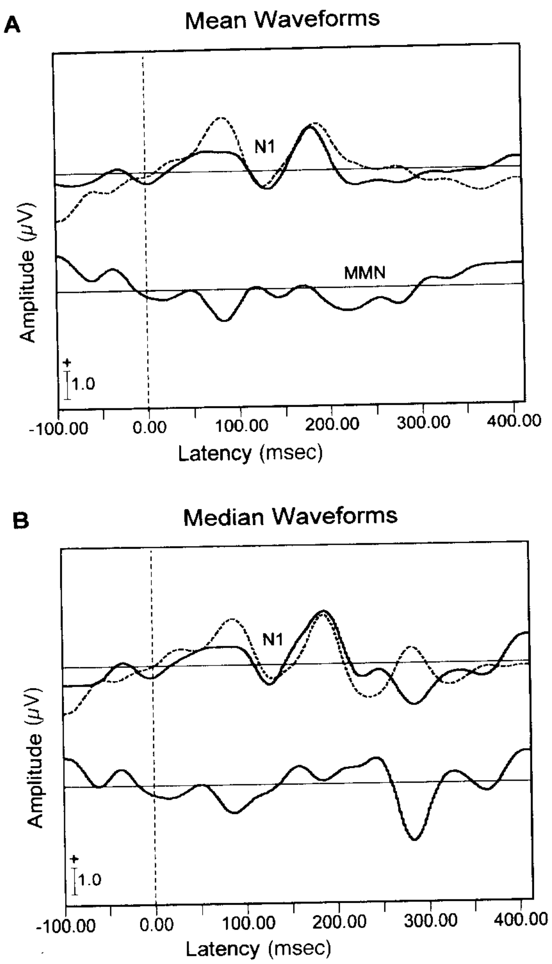

Figure 3 Standard, deviant, and difference waveforms computed using (*A*) the averaging method and (*B*) the median method. The standard (*dotted line*) and deviant (*solid line*) waveforms are overlaid with the difference waveform positioned directly below. *A*, Note the negative deflection in the mean difference waveform that meets both criteria used to validate the MMN (i.e., area $\geq$ 29 msec $\times$ μV; onset latency $\leq$ 217 msec). Because this record was obtained in a control (i.e., no contrast) condition, the "MMN" identified in the mean waveform is a false alarm. *B*, The negative deflection in the difference waveform does not meet the onset latency criterion used in this study; thus, a correct rejection is achieved.

skewed. To illustrate, sample deviant waveforms from one record in Data Set 1 are displayed in Figure 2. Because these data were elicited in a contrast condition, they should contain the signal designated as the MMN. Mean and median waveforms would be expected to look more similar during the time frames associated with components like the N1 and the MMN (relative to earlier or later segments of the epoch). That is, the variability of corresponding distributions should be reduced by the presence of a signal that is time-locked to the stimulus. However, this does not appear to be the case for the waveforms shown in Figure 2. Divergence of mean and median waveforms is observed throughout the epoch. The single explanation for the difference between mean and median values is asymmetry of the underlying voltage distribution.

An asymmetric distribution may result from the presence of extreme values; when these values belong to a different distribution, they are considered contaminants (Barnet and Lewis, 1984). For example, an unrejected artifact in the ERP distribution (e.g., eye blink) is a contaminant because its voltage reflects noise rather than the signal. Whereas the mean is biased by these extreme values, the median is robust to contaminants of relatively large magnitude. Thus, observed asymmetry of the MMN distribution (as implied in Fig. 2) suggests that the signal is more accurately represented by the median values.

As a result, it could be argued that the "true" evoked response (or lack of one) is revealed in median waveforms. Indeed, use of the median instead of the mean has been shown to alter the outcome (i.e., whether or not an MMN is identified) obtained for a given record. Figure 3 displays mean and median (standard, deviant, and difference) waveforms computed using data recorded in a control condition. Because no stimulus contrast was present to elicit the MMN, the "response" validated in the mean difference waveform (i.e., a false alarm) does not reflect a genuine ERP; rather, the negative deflection could only have been created by residual noise in the average. Lack of an analogous response in the median waveforms (i.e., a correct rejection) indicates that median values were more accurate in this particular case.

The poor rate of MMN identification obtained for median waveforms in the present investigation indicates that the median method was unable to overcome an insufficient SNR in the majority of cases. Presumably, the median is susceptible to contaminants in the voltage distribution that are not extreme values. Nonetheless, the theoretical basis for using the median should (at the very least) be considered when interpreting MMN data. That is, outcome (as defined herein) could be dependent on accurate representation of the voltage distributions underlying the ERP waveform. Researchers must guard against continued use of signal averaging based simply on "the inertia of precedent, rather than a thoughtful examination of what information is useful" (Glass and Hopkins, 1996, p. 61).

Acknowledgment. The authors wish to thank Dr. Robert A. Fox for creating the stimulus continuum, Dr. Randall R. Robey for statistical consultation, and the Graham and Sharon Adelman family for the funding of laboratory equipment.

REFERENCES

Alho K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear* 16:38–51.

Barnet V, Lewis T. (1984). *Outliers in Statistical Data*. 2nd Ed. Chichester, UK: John Wiley & Sons.

Borda RP, Frost JD. (1968). Error reduction in small sample averaging through the use of the median rather than the mean. *Electroencephalogr Clin Neurophysiol* 25:391–392.

Chertoff ME, Goldstein R, Mease MR. (1988). Early event-related potentials with passive subject participation. *J Speech Hear Res* 31:460–465.

Dalebout SD, Fox LG. (2000). Identification of the mismatch negativity in the responses of individual listeners. *J Am Acad Audiol* 11:12–22.

Dalebout SD, Fox LG. (2001). Reliability of the mismatch negativity in the responses of individual listeners. *J Am Acad Audiol* 12:245–253.

Dalebout SD, Stack JW. (1999). Mismatch negativity to acoustic differences not differentiated behaviorally. *J Am Acad Audiol* 7:388–399.

Glass GV, Hopkins KD. (1996). *Statistical Methods in Education and Psychology*. 3rd Ed. Boston: Allyn and Bacon.

Hall JW III. (1992). *Handbook of Auditory Evoked Responses*. Boston: Allyn and Bacon.

Hyde ML. (1994). Signal processing and analysis. In: Jacobson J, ed. *Principles & Applications in Auditory Evoked Potentials*. Boston: Allyn and Bacon, 47–83.

Kraus N, Cheour M. (2000). Speech sound representation in the brain. *Audiol Neurootol* 5:140–150.

Kraus N, McGee T. (1994). Auditory event-related potentials. In: Katz J, ed. *Handbook of Clinical Audiology*. 4th Ed. Baltimore: Williams & Wilkins, 406–423.

McGee T, Kraus N, Nicol T. (1997). Is it really a mismatch negativity? As assessment of methods for determining response validity in individual subjects. *Electroencephalogr Clin Neurophysiol* 10:359–368.

Picton TW. (1995). The neurophysiological evaluation of auditory discrimination. *Ear Hear* 16:1–5.

Picton TW, Alain C, Otten L, et al. (2000). Mismatch negativity: different water in the same river. *Audiol Neurootol* 5:111–139.

Ponton C, Don M, Eggermont JJ, Kwong B. (1997). Integrated mismatch negativity (MMNi): a noise free representation of evoked responses allowing single point distribution-free statistical tests. *Electroencephalogr Clin Neurophysiol* 104:381–382.

Schröger E. (1998). Measurement and interpretation of the mismatch negativity. *Behav Res Methods Instrum Comput* 30:131–145.

Sinkkonen J, Tervaniemi M. (2000). Towards optimal recording and analysis of the mismatch negativity. *Audiol Neuroootol* 5:235–246.

Tremblay K, Kraus N, McGee T, et al. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear* 22:79–90.

Tucker DA, Dietrich S, McPherson DL, Salamat MT. (2001). Effect of stimulus intensity level on auditory middle latency response brain maps in human adults. *J Am Acad Audiol* 12:223–232.

Yabe H. (1997). Median method for eliminating infrequent artifacts and identifying the signals blurred by latency jitter and uncertain occurrence. *Behav Res Methods Instrum Comput* 30:68–77.

Yabe H, Saito F, Fukushima Y. (1993). Median method for detecting endogenous event-related brain potentials. *Electroencephalogr Clin Neurophysiol* 89:403–407.