Reliability of the Mismatch Negativity in the Responses of Individual Listeners

Susan D. Dalebout*
Lisa G. Fox*

Abstract

The mismatch negativity (MMN) was recorded from 12 normal adults during four biweekly sessions. Responses were elicited by a synthetically generated speech contrast (/dα/-/gα/) that all listeners discriminated with at least 90 percent accuracy. Standard and deviant waveforms were replicable across sessions for all listeners; however, replicability of the derived difference waveforms was poor. Of greater importance, the MMN identification rate was too low (29%) to allow reliability to be evaluated. The implications that these findings may have on clinical applicability are discussed.

Key Words: Auditory evoked potentials, listening training, mismatch negativity, reliability, replicability, stability

Abbreviations: ALR = auditory late response, MMN = mismatch negativity, VOT = voice onset time

The mismatch negativity (MMN) is an event-related potential that reflects the preconscious detection of change within a stream of acoustic events. It is elicited using an oddball paradigm in which a train of repeating "standard" stimuli is intermittently interrupted by a "deviant" (oddball) stimulus. The paradigm is based on Näätänen's (1995) contention that a neural trace or template is formed to represent the standard stimulus and held in short-term memory. The event-related (MMN) response occurs when a mismatch between this template and the neural representation of the deviant stimulus is detected by mechanisms within the central auditory system. Thus, both stimuli elicit the obligatory components of the auditory late response (ALR) (P_1, N_1, P_2), but the MMN is generated only in response to deviant stimuli. Separately averaging single responses to standard stimuli (i.e., the "standard waveform") and single responses to deviant stimuli (i.e., the "deviant waveform") allows the MMN to be visualized as a negative deflection in the deviant waveform relative to the standard waveform.

Although the MMN is used only for research purposes at the present time, there is considerable excitement about its potential as a clinical tool. At least two features make it attractive for clinical use. First, it is passively elicited, occurring without attention or response from the listener. As a result, the MMN could be useful in the assessment of individuals unable or unwilling to participate in conventional test protocols. Second, its presence indicates that the deviant stimulus has been differentiated from the standard stimulus, suggesting that it may have utility as an objective measure of auditory discrimination ability (Näätänen et al, 1978; Kraus et al, 1993, 1996).

A number of clinical applications have been proposed. For example, the MMN might be used to differentiate individuals with disorders of auditory perception from those with disorders originating at higher levels of function (e.g., involving language, attention, or memory). The MMN also holds promise as a tool for determining the effectiveness of auditory training in individuals with hearing loss or auditory perceptual deficits. In fact, it already has been used to demonstrate the effects of listening training

*Communication Disorders Program, University of Virginia, Charlottesville, Virginia

Reprint requests: Susan Dalebout, Communication Disorders Program, University of Virginia, 2205 Fontaine Avenue, Suite 202, Charlottesville, VA 22903

in several research investigations (Kraus et al, 1995; Tremblay et al, 1997, 1998).

In one such study, Kraus et al (1995) sought to determine if training listeners to discriminate a difficult speech contrast (i.e., two variants of a synthetic speech token /dα/) would change the neurophysiology of the central auditory system. As expected, training improved the listeners' ability to behaviorally discriminate the speech stimuli. More interesting, however, was the finding that mean MMN duration and magnitude were also enhanced. An MMN was considered to be present in 10 of 13 listeners before training and all listeners after training. In addition, an increase in MMN duration was observed in 12 of 13 listeners and a considerable increase in MMN area was observed in 11 of 13 listeners. It was concluded that the ability to demonstrate changes in MMN characteristics following speech perception training makes it possible to objectively evaluate the efficacy of auditory rehabilitation strategies.

Tremblay et al (1997) also reported that behavioral training can modify the neurophysiologic representations of speech sounds. In addition, the hypothesis that training-induced changes can generalize to untrained stimuli was tested. More specifically, the study was designed to determine if transfer of learning would be evident behaviorally and neurophysiologically in response to a novel speech contrast that shared the same voice onset time (VOT) as the trained contrast but had a different place of articulation. Both contrasts were based on cues that are not phonemic in English. As predicted, training improved the ability of listeners to discriminate both the trained and untrained contrasts. There was no significant intersession change in behavioral performance for listeners in a control group who did not receive training. Similarly, the composite MMN for the experimental group was enhanced in response to both contrasts following training, but no significant intersession changes were observed for the control group. Tremblay et al considered these findings important from a clinical perspective, suggesting that they represent an objective technique for examining the neurophysiologic changes that may result from (re)habilitative efforts with children who have hearing loss or auditory-based learning problems, some of whom may not be able to participate in behavioral testing. Further, the authors conceptualized the MMN as a means for exploring the human capacity for change and for understanding processes associated with foreign language training, musician ear training, and other forms of auditory learning.

In a subsequent study, Tremblay et al (1998) investigated the time course of changes in the MMN and behavioral performance as a function of listening training. Listeners were trained to discriminate and identify synthetic speech stimuli differing in VOT. Once again, mean behavioral performance was improved, and the group MMN was enhanced following training. Moreover, it was reported that 9 of 10 listeners learned to identify the novel stimuli, and all 10 listeners showed significant changes in the MMN. Training-associated changes in the MMN always preceded or occurred on days concurrent with changes in behavioral performance. Tremblay et al concluded that neurophysiologic changes observed during auditory training indicate that the training program is effectively altering neural representations of speech sounds and that changes in behavior are likely to follow. A change in neurophysiology that is *not* accompanied by behavioral change implies that the intervention method has successfully altered the brain's ability to code speech sounds but that behavioral changes have been slowed or prevented by nonauditory problems involving cognition, motivation, etc.

The results of these studies suggest that the MMN is capable of reflecting the physiologic plasticity associated with auditory learning. However, before the MMN can be used as a clinical tool, it will be necessary to demonstrate that it is (a) detectable in virtually all neurologically normal individuals able to differentiate the eliciting stimuli and (b) replicable within individuals over successive sessions. Recently, Dalebout and Fox (2000) reported that the MMN was not detectable in the responses of all individuals able to differentiate the eliciting stimuli. In that study, the MMN could be identified in only 33 percent of the responses recorded from listeners with behavioral discrimination scores greater than 90 percent. In this follow-up report, attention turns to the reliability of the MMN in individual listeners over repeated sessions.

Intersession reliability of the MMN has been examined in a number of studies (Chertoff et al, 1988; Lang et al, 1995; Pekkonen et al, 1995; Escera and Grau, 1996; Frodl-Bauch et al, 1997; Deouell and Bentin, 1998; Joutsiniemi et al, 1998; Kathmann et al, 1999; Tervaniemi et al, 1999; Escera et al, 2000). However, generalizing the findings of these studies to the present one is difficult for several reasons.

First, nonspeech signals (e.g., tone bursts) were used as stimuli in the preceding investigations. The MMN elicited by speech stimuli (as in the present study) is likely to differ from the MMN elicited by simpler stimuli along a number of dimensions, making generalization inappropriate.

Second, none of these studies used techniques for validating the presence of an MMN (i.e., for maximizing the probability that negative voltage recorded within a designated time window is indeed an MMN). For example, studies incorporating a control (i.e., no contrast) condition (McGee et al, 1997; Dalebout and Fox, 2000) have demonstrated that some negative components visually identified as MMN responses are actually false positives (i.e., MMNs identified in responses recorded in a control condition). By applying signal detection techniques to data analysis, validation criteria can be devised that maximize the probability of identifying true MMN responses. A variety of different approaches to response validation have been explored (McGee et al, 1997).

Third, the results of some studies suggest that MMN reliability differs as a function of stimulus and recording parameters. For example, Escera and Grau (1996) reported significant correlation coefficients at two of six electrode sites for MMN amplitude and one of six electrode sites for MMN latency. Similarly, Deouell and Bentin (1998) found significant intersession correlations for amplitude and latency when one of four stimulus contrasts was used. Pekkonen et al (1995) reported a significant correlation for MMN amplitude at one of seven electrode sites only when using one of two stimulus contrasts and one of two interstimulus intervals. It does not appear that an alpha-level adjustment was used to compensate for the large number of statistical tests performed in these studies. Because the use of multiple statistical procedures inflates the familywise type I error rate (i.e., the probability of erroneously rejecting at least one null hypothesis in a set of statistical tests) and can result in spurious findings, it is possible that the correlation coefficients reported in the aforementioned studies might have been due to chance.

It is widely accepted that an instrument cannot be considered valid unless it is reliable (Krathwhol, 1998). Evidence demonstrating reliability of the MMN in individual listeners is neither convincing nor complete. Therefore, the present study was undertaken to evaluate the test–retest reliability of MMN recorded from normal adults using an easily discernible speech contrast.

METHOD

Stimuli

Standard and deviant stimuli were taken from a nine-item synthetic speech continuum that varied in place of articulation from /dα/ to /gα/. Stimulus parameters have been described in greater detail elsewhere (Dalebout and Stack, 1999; Dalebout and Fox, 2000). Syllables on the continuum differed only in the starting frequency of their second and third formant (F2, F3) transitions. The F2 starting points varied in equal steps from 1700 (endpoint of /dα/) to 2100 Hz (endpoint of /gα/). The starting frequencies of F3 varied in equal steps from 2800 (endpoint of /dα/) to 2100 Hz (endpoint of /gα/). Transition durations for F2 and F3 were 40 msec and total stimulus duration was 90 msec.

It was important for the eliciting stimulus contrast to be easily discriminated so that test–retest variance in the MMN could be isolated from a learning effect associated with repeated testing. Thus, the continuum endpoints (i.e., steps 1 and 9) formed the stimulus contrast.

Listeners

Twelve graduate students served as listeners in the study (mean age = 23 years; range = 20–26 years). All listeners were in good health with a negative history of learning disabilities, neurologic disorders, and significant head trauma (defined as head trauma resulting in unconsciousness for a period of 2 minutes or longer). In addition, each listener had normal hearing and middle-ear function bilaterally, as determined by pure-tone screening (20 dB HL at 500–8000 Hz) and tympanometry.

All listeners were able to differentiate the eliciting stimulus contrast with at least 90 percent accuracy as measured in a two-alternative, forced-choice, same/different discrimination task. Stimuli were presented in pairs, and listeners were asked to decide if the tokens in each pair were the same token presented twice or two different tokens. Listeners marked "same" or "different" on an answer sheet. A total of 140 "same" (1–1, 9–9) and "different" (1–9) pairs were presented.

Electrophysiologic Procedures

Electrophysiologic data were collected (using Neuroscan hardware and software) in four identical recording sessions occurring at 2-week intervals. In an attempt to reduce the variability associated with differences in listener state (i.e., level of arousal), each listener's sessions were scheduled at the same time of day and (with a few exceptions) on the same day of the week. Tympanometry was performed at the beginning of each session to ensure continued normal middle-ear function.

Listeners reclined in a sound-attenuating, electromagnetically shielded chamber and watched videotaped movies of their own choosing. Videotape audio levels averaged 40 dB SPL (A-weighted scale), measured at the listener's left ear. Experimental stimuli were presented to the right ear at 72 dB SPL through an Etymotic ER-3A insert earphone. The interstimulus interval was 1.1 seconds and the probability ratio between standard and deviant stimuli was 85/15. Evoked responses were recorded at Fz, Cz, and both mastoids, with the nose as reference and the forehead as ground. For this study, the listener's optimal response (recorded at either Fz or Cz) was used for analysis. Eye artifact was recorded with a bipolar electrode montage, using supraorbital and infraorbital electrodes around the left eye. The recording window was 550 msec (including a 100-msec prestimulus period). Evoked responses were analog filtered online from 0.1 to 100 Hz, a bandpass setting sufficiently wide to preclude the distortion of waveform components.

Responses were also recorded in a control condition using different but comparable listeners (Dalebout and Fox, 2000). Recording parameters were identical to those used in the contrast condition with one exception: trial blocks consisted of 1000 events in the control condition and 500 events in the contrast condition. However, the total number of responses recorded from each listener in each condition was the same (i.e., 1800 after online rejection of activity exceeding $\pm$ 100 μV).

Data Analysis Procedures

Responses to standard and deviant stimuli were averaged separately; responses to standard stimuli following deviants were excluded from standard response averages. Averaged waveforms were digitally low-pass filtered at 30 Hz to further remove activity outside the frequency range of the MMN. Linear trends were removed using a software routine provided by Neuroscan, Inc. Detrending corrects for a form of artifact initiated when sudden responses cause the analog filter to "ring," thus producing a large, slow recovery waveform that extends beyond the sampling epoch. (Examination of waveforms before and after detrending showed that this correction did not alter outcome in any way.) Finally, each listener's averaged response to standard stimuli was subtracted from the corresponding deviant response to create a difference waveform for each session. The same process was followed in preparing data from the control condition in which standard and deviant stimuli were identical.

The MMN was visually identified as a negative deflection in the difference waveform following N_1 onset in the standard and deviant waveforms. Visually identified negative components were validated as MMN responses using an area criterion of 110 msec $\times$ μV. This validation technique was based on the report of McGee et al (1997), who compared 17 combinations of techniques and criteria for validating the MMN in individual responses. These investigators concluded that establishment of threshold criteria for waveform characteristics (e.g., area, onset latency, duration) produces the best combination of MMN hit rate and d′ value. d′ is a statistic that incorporates both "hit rate" (i.e., the percentage of presumably correct identifications of the MMN in a contrast condition) and "false alarm rate" (i.e., the percentage of incorrect identifications of the MMN in a control condition). More accurate tests are associated with larger d′ values (Swets, 1964). As in the McGee et al (1997) study, the specific criterion value used in the present study was established by applying signal detection techniques to data analysis. First, examiners blind to stimulus condition visually identified the MMNs in coded responses that had been recorded in both contrast (session 1 data) and control conditions. Responses were identified by condition only after decisions regarding the presence or absence of an MMN had been made. Second, cutoff criteria for an MMN area and onset latency were systematically varied until the optimal combination of hit rate and false alarm rate was determined on the basis of d′. In this particular data set, the largest d′ value was obtained using a threshold criterion for a single waveform characteristic (i.e., area exceeding 110 msec $\times$ μV). In the McGee et al study, the combination of area and onset latency criteria was the most accurate, whereas area was the most useful single measure. In the present study, all validated MMN responses had onset latencies of 205 msec or less.

MMN onset and offset were defined as the peaks preceding and following the visually identified negative component in the difference waveform. Selection of these peaks was based on an examination of difference waveforms, as well as the standard and deviant waveforms from which they were derived. Area was measured by drawing a line between the onset and offset points and calculating the enclosed area in msec × µV (Kraus et al, 1995; Tremblay et al, 1997, 1998). In this study, data were converted to ASCII format and exported to a spreadsheet for area computation.

RESULTS

Replicability of the Auditory Late Response

Components of the ALR (P_1, N_1, P_2) were highly replicable across sessions for all listeners in the study. This is represented in Figure 1, where standard and deviant waveforms from all four sessions have been overlaid for each of four listeners (listeners A–D). The listeners were chosen to represent the range of MMNs identified per listener across the four sessions (range = 0–3). That is, listeners A, B, C, and D demonstrated 0, 1, 2, and 3 MMN responses, respectively, over the four sessions. Detectability of the MMN in the difference waveforms did not appear to be related to the replicability of the standard and deviant waveforms from which they were derived. For example, waveforms recorded from listener B (with one identifiable MMN response among the four sessions) were as replicable as those of listener D (with the greatest number of identifiable MMN responses among the four sessions).

Notably, the morphology of standard and deviant waveforms was distinctive for each listener. That is, waveforms obtained from the same individual were more similar to one another than to those of other listeners. In

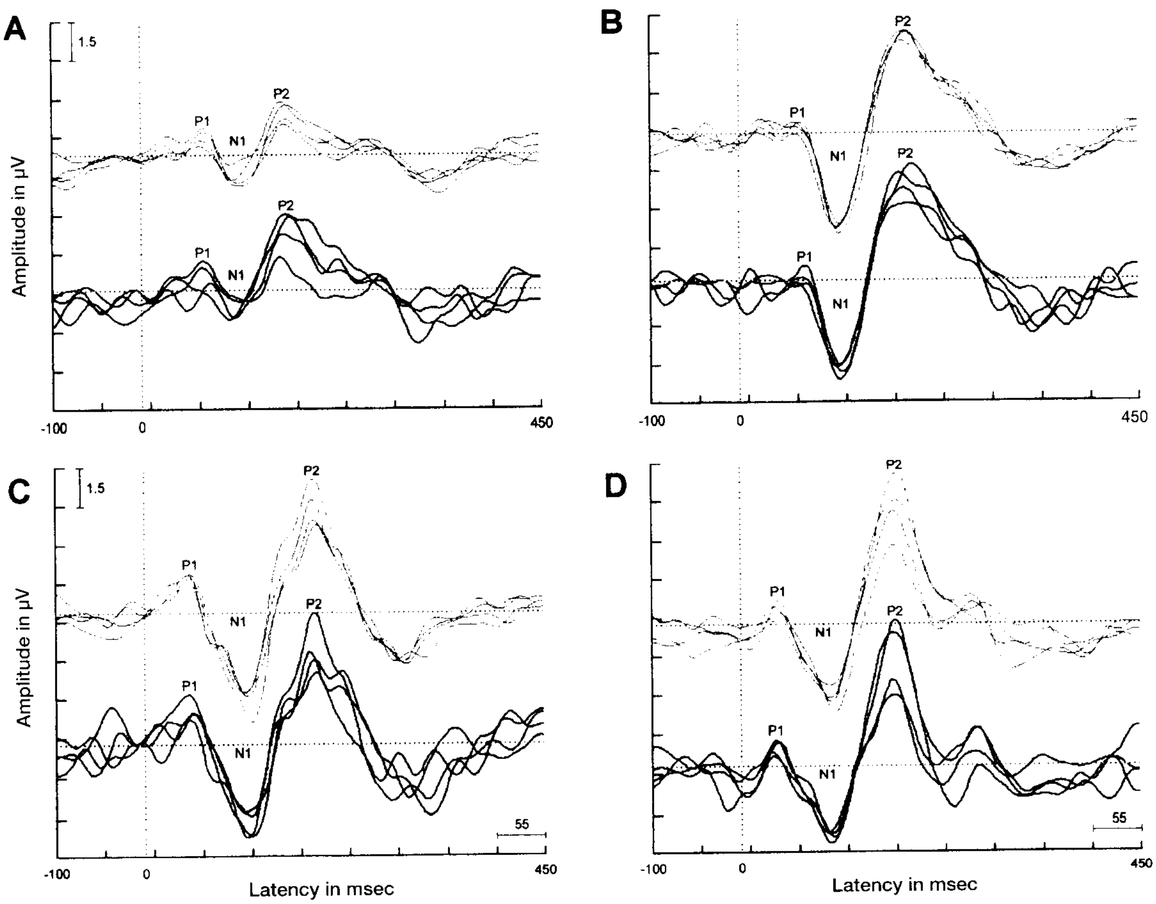

Figure 1 Standard and deviant waveforms recorded from four listeners (A–D) in four sessions (1–4) are overlaid to demonstrate replicability of the auditory late response. Listeners were chosen to represent the range of MMN responses identified per listener (0–3). In this and subsequent figures, standard waveforms are shown in gray and deviant waveforms in black, positive polarity is displayed as up-going, and waveforms are from Cz.

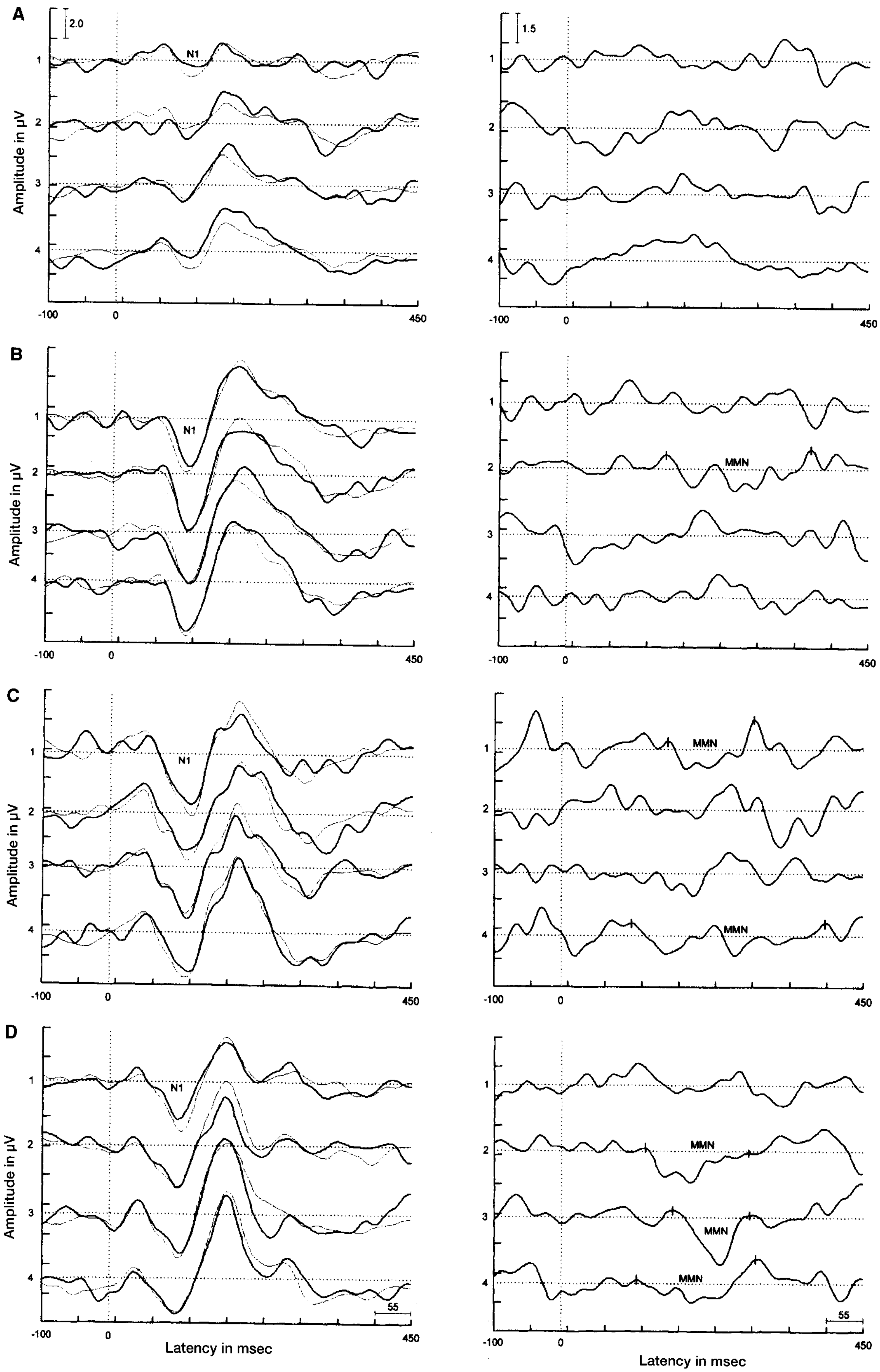
A
B
C
D
2.0
1.5
N1
MMN
Amplitude in µV
1
2
3
4
-100
0
450
55
Latency in msec

Figure 2 Panels on the left display standard and deviant waveforms from four sessions (1–4) for the same four listeners (A–D) depicted in Figure 1. Panels on the right show difference waveforms derived from the corresponding standard and deviant waveforms on the left. When present, an MMN is shown as the (shaded) difference between the standard and deviant waveforms (when the deviant waveform is more negative); vertical hatchmarks denote MMN onset and offset in the difference waveforms. Listener A failed to demonstrate identifiable MMNs in any session. Listener B demonstrated MMNs in session 2. Listener C demonstrated MMNs in sessions 1 and 4. Listener D demonstrated MMNs in sessions 2, 3, and 4. Although standard and deviant waveforms are replicable within individual listeners across sessions, difference waveforms are not.

general, the latency of ALR components appeared more stable than their amplitude.

Replicability of Difference Waveforms

Although standard and deviant waveforms were replicable, there was considerable variation among the four difference waveforms derived for each listener. This was true for all listeners in the study. Figure 2 displays the standard and deviant waveforms (left panels) and corresponding difference waveforms (right panels) for the four listeners depicted in Figure 1. Even among the few listeners for whom an MMN was identified on more than one occasion (e.g., listeners C and D), replicability of the difference waveforms, and therefore the MMN, was poor. MMN responses varied in morphology, onset and offset latency, magnitude, and duration.

Incidence of Mismatch Negativity

Forty-eight difference waveforms were examined for the presence of an MMN (12 listeners × 4 sessions = 48). The rate of MMN identification was extremely low. Only 14 of the 48 waveforms (29%) had negative components that met the validation criterion used in this study. As shown in Figure 3, the number of MMNs identified among the four responses recorded from each listener ranged from zero to three; none of the listeners had an identifiable response in all four sessions, and two failed to demonstrate an MMN in any session. The distribution of MMN responses identified across the four sessions is shown in Figure 4. Importantly, the number of MMN responses identified in the first and last sessions was the same.

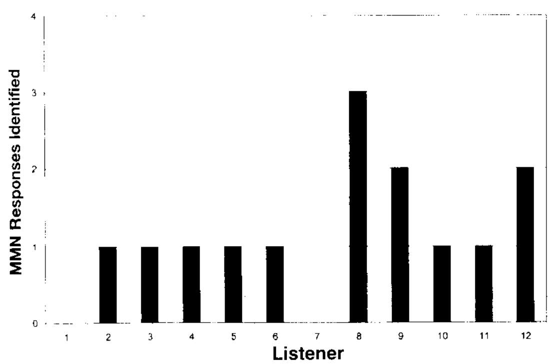

Figure 3 Number of MMN responses identified for each listener over four recording sessions.

DISCUSSION

Ordinarily, when an event-related potential serves as the dependent variable in an experiment, variation in a waveform characteristic (e.g., area, onset latency) is attributed to listener trait variables (e.g., age, gender, patient diagnosis) or experimentally manipulated variables of interest (e.g., listening training, task difficulty). However, additional sources of variance exist that are neither manipulated nor measured by the examiner (Segalowitz and Barnes, 1993). For example, the listener's state (i.e., arousal level) is likely to vary from one trial (or session) to the next, without experimental manipulation. Another source of uncontrolled variance is the inherent unreliability of the event-related potential itself, that is, instability

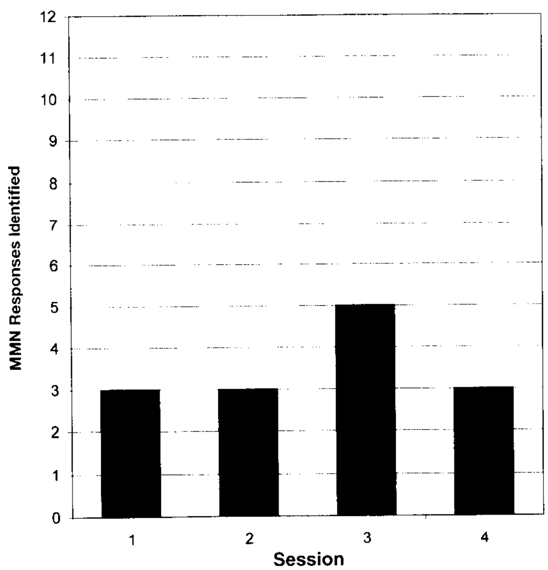

Figure 4 Distribution of MMN responses identified across four recording sessions.

in the generation of a specific response and imprecision in its identification and measurement. The expectation is that the strength of the manipulated variance will overcome the uncontrolled variance associated with fluctuations in listener state and measurement error to reveal the relationship of interest (Segalowitz and Barnes, 1993). However, the variance associated with measurement error and fluctuation in listener state is critical because it impacts the validity of the experimental effect.

The objective of the present investigation was to evaluate the reliability of the speech-elicited MMN without manipulation of listener trait variables or experimental conditions. With these held constant, the variability observed theoretically represents the uncontrolled variance attributable to fluctuations in listener state and measurement error. Unfortunately, observation of uncontrolled variance in the MMN was precluded by the low identification rate. Nevertheless, at least three conclusions can be drawn from these findings.

First, standard and deviant waveforms were replicable across sessions for individual listeners and distinctive from those of other listeners. In other words, listeners looked more like themselves than one another. This is consistent with the findings of Dalebout and Robey (1997), who also reported that multiple standard and deviant waveforms from the same listener were remarkably similar yet unique when compared with those of other listeners.

Second, although standard and deviant waveforms were replicable across sessions, slight differences attributable to normal test–retest variation are to be expected and were observed. Of critical importance, however, is the fact that standard and deviant waveforms varied *independently* of one another, resulting in highly variable difference waveforms. In fact, difference waveforms for the same listener were no more alike than those compared across listeners. Even when an MMN was identified on more than one occasion in the same listener, difference waveforms (and MMN components) did not replicate.

Third, an equal number of MMN responses identified in the first and last sessions suggests that MMN detectability was neither enhanced nor diminished by repeated exposure to the stimuli. Moreover, post hoc analysis of the data from six listeners (24 records), in which (a) single responses were averaged in smaller blocks (block = 100 responses) and (b) blocks were averaged cumulatively (i.e., average of block 1 + block 2, average of block 1 + block 2 + block 3, etc.), allowed the development of each record to be evaluated. No evidence of habituation, fatigue, or learning during the course of a session was observed.

Unfortunately, MMN reliability could not be adequately evaluated in this study due to the poor rate of MMN identification (29%). Although the MMN is reported to be robust at the group level, it is often difficult to identify it in the responses of individual listeners (Kurtzberg et al, 1995; Lang et al, 1995; McGee et al, 1997; Ponton et al, 1997; Dalebout and Fox, 2000). Therefore, methods for enhancing MMN detectability should be explored. For example, the finding that derived difference waveforms were not replicable in individual listeners suggests that this method of identifying the MMN may not be ideal. Development of alternative techniques for quantifying differences between standard and deviant waveforms would avoid the additional variability introduced by the computation of difference waveforms.

Acknowledgment. The authors wish to thank Kelly Mankin for her assistance with data collection, Kelli Pugh for her assistance with post hoc data analyses, the audiology graduate students who participated as listeners in this study, and the Graham and Sharon Adelman family for their funding of laboratory equipment.

REFERENCES

Chertoff M, Goldstein R, Mease M. (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, Robey RR. (1997). Comparison of the intersubject and intrasubject variability of exogenous and endogenous auditory evoked potentials. *J Am Acad Audiol* 8:342–354.

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

Deouell LY, Bentin S. (1998). Variable cerebral responses to equally distinct deviance in four auditory dimensions: a mismatch negativity study. *Psychophysiology* 35:745–754.

Escera C, Grau C. (1996). Short-term replicability of the mismatch negativity. *Electroencephalogr Clin Neurophysiol* 100:549–554.

Escara C, Yago E, Polo MD, Grau C. (2000). The individual replicability of mismatch negativity at short and long inter-stimulus intervals. *Clin Neurophysiol* 111:546–551.

Frodl-Bauch T, Kathmann N, Moller H, Hegerl U. (1997). Dipole localization and test–retest reliability of frequency and duration mismatch negativity generator processes. *Brain Topogr* 10:3–8.

Joutsiniemi SL, Ilvonen T, Sinkkonen J, Huotilainen M, Tervaniemi M, Lehtokoski A, Rinne T, Näätänen R. (1998). The mismatch negativity for duration decrement of auditory stimuli in healthy subjects. *Electroencephalogr Clin Neurophysiol* 108:154–159.

Kathmann N, Frodl-Bauch T, Hegerl U. (1999). Stability of the mismatch negativity under different stimulus and attention conditions. *Clin Neurophysiol* 110:317–323.

Krathwhol DR. (1998). *Methods of Educational and Social Science Research: An Integrated Approach*. 2nd Ed. New York: Addison-Wesley.

Kraus N. (1996). The discriminating brain: MMN and acoustic change. *Hear J* 49(5):10–43.

Kraus N, McGee T, Ferre J, Hoeppner J, Carrell T, Sharma A, Nicol T. (1993). Mismatch negativity in the neurophysiologic/behavioral evaluation of auditory processing deficits: a case study. *Ear Hear* 14:223–234.

Kraus N, McGee T, Carrell T, King C, Tremblay K, Nicol T. (1995). Central auditory system plasticity associated with speech discrimination training. *J Cogn Neurosci* 7:25–32.

Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 273:971–973.

Kurtzberg D, Vaughan HG, Kreuzer JA, Fliegler KZ. (1995). Developmental studies and clinical application of mismatch negativity: problems and prospects. *Ear Hear* 16:104–117.

Lang AH, Eerola O, Korpilahiti P, Holopainen I, Salo S, Aaltonen O. (1995). Practical issues in the clinical application of mismatch negativity. *Ear Hear* 16:118–130.

McGee T, Kraus N, Nicol T. (1997). Is it really a mismatch negativity? An assessment of methods for determining response validity in individual subjects. *Electroencephalogr Clin Neurophysiol* 104:359–368.

Näätänen R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear* 16:6–18.

Näätänen R, Gaillard AWK, Mantysalo S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol* 42:313–329.

Pekkonen E, Rinne R, Näätänen R. (1995). Variability and replicability of the mismatch negativity. *Electroencephalogr Clin Neurophysiol* 96:546–554.

Ponton CW, Don M, Eggermont JJ, Kwong B. (1997). Integrated mismatch negativity (MMN): a noise-free representation of evoked responses allowing single-point distribution-free statistical tests. *Electroencephalogr Clin Neurophysiol* 104:143–150.

Segalowitz SJ, Barnes KL. (1993). The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology* 30:451–459.

Swets JA. (1964). *Signal Detection and Recognition by Human Observers*. New York: John Wiley & Sons.

Tervaniemi M, Lehtokoski A, Sinkkonen J, Virtanen J, Ilmoniemi RJ, Näätänen R. (1999). Test–retest reliability of mismatch negativity for duration, frequency, and intensity changes. *Clin Neurophysiol* 110:1388–1393.

Tremblay K, Kraus N, Carrell TD, McGee T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J Acoust Soc Am* 102:3762–3773.

Tremblay K, Kraus N, McGee T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport* 9:3557–3560.