ELSEVIER

Neuroscience Letters 341 (2003) 251–255

Neuroscience Letters

www.elsevier.com/locate/neulet

Quantifying signal-to-noise ratio of mismatch negativity in humans

Anthony T. Cacace[a,*], Dennis J. McFarland[b]

[a]Department of Surgery, Albany Medical College, Albany, New York, NY 12208, USA
[b]Laboratory of Nervous System Disorders, The Wadsworth Center, New York State Health Department, Albany, New York, NY 12201, USA

Received 1 November 2002; received in revised form 14 January 2003; accepted 29 January 2003

Abstract

Mismatch negativity (MMN) is thought to represent a neurophysiological index of auditory information processing that is independent of attention. Because this measure does not require an overt behavioral response, MMN has potential to evaluate higher order perceptual abilities in infants, young children and difficult-to-test populations, thereby extending results obtained from more basic physiologic and electroacoustic measures (auditory brainstem responses, ABRs; otoacoustic emissions, OAEs). Whereas the basic tenet of MMN is appealing, several issues-of-contention remain to be solved before this event related potential (ERP) can be applicable for routine clinical use. These issues include the consistent identification of MMN within individuals (vs. groups), its stability over time, and its reportedly poor signal-to-noise ratio (SNR). Herein, we focus on the issue of SNR, by comparing and contrasting SNR of MMN with other long latency auditory ERPs.

Keywords: Mismatch negativity; Auditory brainstem responses; Event related potential; Otoacoustic emissions; Signal-to-noise ratio

Mismatch negativity (MMN) is thought to represent a neurophysiological index of auditory information processing that is independent of attention [15]. It is elicited within a passive oddity paradigm whereby binary sequential stimuli consisting of infrequent deviants (oddballs) are presented randomly within a stream of frequently occurring standards. If the deviant stimulus is perceptually different from the standard, then MMN is manifest as an enhanced negativity; quantified as a difference wave between standard and deviant time averaged waveforms [15]. Onset latency is in the 100–200 ms range and the response overlaps with well-known long latency auditory event related potentials (ERPs) in this time frame and beyond.

By engaging brain systems involved in processing stimulus deviancy, MMN is considered to be an index of discrimination ability. Indeed, the potential to assess sensory discrimination (perceptual ability) independent of attention can be particularly valuable in the assessment of infants, young children and difficult-to-test populations, where perceptual skills cannot be obtained easily by conventional behavioral methods. Whereas many relevant studies have been performed on this topic, the downside of this measurement concerns the consistency by which MMN can be identified within individuals (vs. groups) and the extent to which MMN is a stable and reliable indicator over time [4,6,13,17]. Both factors are related either directly or indirectly to signal-to-noise ratio (SNR). Additionally, many authors have recognized problems associated with this derived response and work aimed at improving SNR has been advocated before MMN can evolve as a viable clinical tool [8–11,14].

Herein, we report on a statistical method for quantifying SNR based on Pearson's product moment correlation coefficient (Pearson's r). We use this quantity as a metric to compare and contrast SNR of MMN with other auditory ERPs (N100 and P300) in this time frame. Because Pearson's r represents a normalized difference metric that is signed, and since it is scale independent, direct comparisons between MMN and other long latency ERPs are possible. This analysis provides a perspective on the robustness of MMN and demonstrates a way of evaluating parameters for further development and study.

Ten adults (four males; six females, ages 18–52 years), with no reported hearing deficits (pure-tone thresholds < 20 dB HL, 0.25–4.0 kHz, bilaterally), participated in this experiment. All individuals were briefed as to the nature of

* Corresponding author. Fax: +1-518-262-6670.
E-mail address: cacacea@mail.amc.edu (A.T. Cacace).

0304-3940/03/$ - see front matter © 2003 Elsevier Science Ireland Ltd. All rights reserved.
doi:10.1016/S0304-3940(03)00196-4

the experiment and provided signed informed consent. The Internal Review Board of the Albany Medical College approved the study. A portion of this paper has been presented in abstract form [1] and represents a subset of responses from a more comprehensive investigation [2]. Whereas the more comprehensive study recorded electroencephalographic (EEG) activity from 24 scalp electrodes, and used various stimuli and task demands, we limit the analysis herein to three central electrodes (Fz, Cz, and Pz) and to pure tone stimulation in an easy discrimination condition. The main intent was to illustrate a novel statistical metric of SNR as a means to facilitate comparison with other auditory ERPs in this time frame.

Electroencephalographic activity was recorded from 24 scalp locations (F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, FPz, Cz, Fz, Pz, A1, A2, M1, M2, PM1, PM2) based on the ten-twenty system of the International Federation [7]. Electrodes were referenced to a sterno-vertebral site and grounded to the left forearm. Two additional vertical and horizontal bipolar channels were used for monitoring electro-ocular (EOC) activity (eye blinks and low frequency lateral eye movements), such that EEG contaminated with high-level EOC activity could be removed from the data prior to analysis. The EEG was acquired using commercially available hardware and software (SCAN and STIM systems, NeuroScan, Inc., El Paso, TX). Additional signal processing (time domain averaging, filtering, statistical analysis, etc.,) was performed off-line.

Stimuli were digitally constructed short duration (50 ms) pure tones, shaped with a Blackman window having 5 ms rise/fall times. The standard or frequent stimulus was always a 250 Hz tone and the target/deviant stimulus was a higher frequency tone, presented at $+6$ just noticeable differences (JNDs) above an individual's differential threshold for 250 Hz using forced-choice psychophysical methods (see Ref. [2] for details). Because deviant/oddball stimuli were constructed individually for each participant, these conditions were considered to be easy discriminations. During data collection, the 250 Hz tone served as the standard stimulus and the higher frequency tones served as the oddball (probability of occurrence; $P = 0.80$ standards, $P = 0.20$ oddballs; total stimulus presentations, 240 trials).

Testing was performed in a lighted, commercially constructed sound-attenuating test booth (Tracoustics, RE 145) where individual participants were seated on a padded reclining chair with head, leg and arm support. Participants were instructed to remain awake, keep eyes open and focus on a designated point on a wall straight ahead. They were also advised to avoid/minimize any unnecessary eye or body movements or muscle contraction during the individual test conditions. Electroencephalographic activity was collected over a 1200 ms time epoch (600 ms pre-stimulus interval; 600 ms post-stimulus interval), amplified $\times 20\,000$, filtered between 0.1 and 300 Hz (12 dB/octave slope; Grass Model 12 Neurodata Acquisition System), digitized at a rate of 1000 Hz with 16-bit resolution and stored in digital form on a trial-by-trial basis. Whereas equal pre and post stimulus intervals were not critical to the SNR computations described herein, this aspect of the experimental design was an essential component to the larger experiment [2], which focused on evaluating event related synchronization and desynchronizations of EEG rhythmicities. Triggering for EEG data acquisition, A/D conversion and stimulus presentation were performed by STIM and SCAN systems. Artifact rejection was used and individual trials of EEG activity exceeding ± 50 μV were excluded from the analysis. Artifacts were monitored on horizontal and vertical electro-ocular channels and at Cz.

Although we were mainly interested in studying MMN using the passive oddball paradigm, active discrimination of oddball stimuli under similar conditions was also included in order to compare and contrast SNR between MMN and the P300 response. Passive and active oddball conditions were performed on separate runs. In the active discrimination condition, individuals were instructed to press a button on an instrument panel (STIM system switch response pad, P/N 1141) with the index finger of their right hand to indicate every time the target/oddball stimulus was discriminated from the stream of standard stimuli. In the passive discrimination condition, individuals were instructed to listen to all stimuli; no button press was required to indicate deviant stimulus identification. Stimuli were presented binaurally at a level approximating 80 dB SPL through insert earphones (Etymotic-ER3A) and pseudo-random interstimulus intervals (3.0 ± 0.5 s) were used.

For both passive and active oddball conditions, responses to frequent, target and deviant stimuli were first sorted and separately averaged in the time domain off-line. Next signal-to-noise ratio was calculated using Pearson's product moment correlation coefficient (Pearson's r), which served as a normalized measure of ERP magnitude. Whereas Pearson's r facilitates comparison of MMN with N100 and P300, it was not used to determine significance between two signals or events. Rather, it was applied as an explicit measure of SNR. Specifically, Pearson's r is the ratio between the cross product of the two signals (numerator) divided by the product of variances of the two signals (denominator). For MMN and P300, the correlation was computed between standard and oddball stimulus conditions. In this computation, the stimulus condition was represented as a dummy variable with zero (0) being the standard and one (1) being the deviant. In this context, 0 signifies not being different from the standard and 1 signifies being quite different from the standard. For N100, the correlation was computed between the silent prestimulus baseline interval and the poststimulus interval. Here, the dummy variable 0 represents the silent prestimulus baseline interval and 1 represents the poststimulus interval, which includes the response to the standard stimulus. Because Pearson's r is a normalized difference metric that is signed (values ranging from -1.0 to $+1.0$) and since this metric is scale independent, direct comparisons between MMN and other long latency ERPs (i.e. N100, P300) are possible.

Pearson's r was computed for each time point in the ERP waveform. Thus, every value of the EEG signal at each point in time was correlated with the corresponding value of the dummy variable representing the stimulus condition. Larger absolute values of the resulting correlations represented higher SNRs. Also examined was the coefficient-of-determination (r^2, the squared value of Pearson's r). This metric allows for comparison of SNR irrespective of the sign and provides a convenient approach to compute the amount of variance in the dependent variable (y) accounted for by the independent variable (x). The r^2 values were used in the statistical comparisons noted below.

The SNR for N100 was computed as the maximum negative voltage within the time interval between 80–160 ms; SNR for P300 was computed as the maximum positive voltage in the time interval between 250 and 500 ms; and, SNR for MMN was computed as the maximum negative voltage in the time interval between 100 and 200 ms. Although time intervals in our analysis were chosen somewhat arbitrarily, they captured the time frame over which responses are known to occur. For each component, the largest response occurring at Fz, Cz, or Pz was used.

Deviant and target stimuli were referenced to the individual's JND performance (average JNDs for 250 Hz tonal stimuli: 13.5 Hz, SD, 5.8 Hz) and represented those binary stimulus conditions that were considered easy to discriminate. Obtaining discrimination thresholds prior to acquiring EEG data and presenting stimuli at +6 JNDs was a central design consideration in this experiment, since it leaves no doubt that stimuli were easily discriminable and that ERPs based on deviant/target detection mechanisms were directly comparable.

Fig. 1 provides an example that compares and contrasts averaged auditory ERP waveforms for active listening (P300 generation; top) and passive listening conditions (MMN generation, middle) from an individual participant. For each time domain waveform designated in the top (standard vs. target) and middle (standard vs. deviant) graphs, SNRs were computed on a point-by-point basis (bottom graph) using the Pearson's r statistic. Maximum correlation coefficients are also presented in numerical form for all participants (Table 1).

Fig. 2 (top) uses bar plots to represent peak latencies of the waveform components; Fig. 2 (bottom) compares SNRs in terms of the coefficient-of-determination. In this context, the r^2 values were used because they provide an index of the absolute size of the SNR (independent of sign) and therefore are more suitable for a statistical comparison of components with opposite sign (i.e. MMN and P300). Furthermore, because r^2 indexes the percent-of-variance accounted for by the correlation coefficient, it represents an equal-interval scale. Using r^2 as the dependent variable, analysis-of-variance (ANOVA) indicated significant differences in SNR between the mean values of r^2 for MMN, N100 and P300 ($F = 4.97$, d.f. $= 2, 18$, $P < 0.0191$). Moreover, marked individual differences in SNR are also apparent for each ERP.

On a comparative basis, SNR of MMN is low with respect to other time-averaged ERPs (N100, P300) (Figs. 1

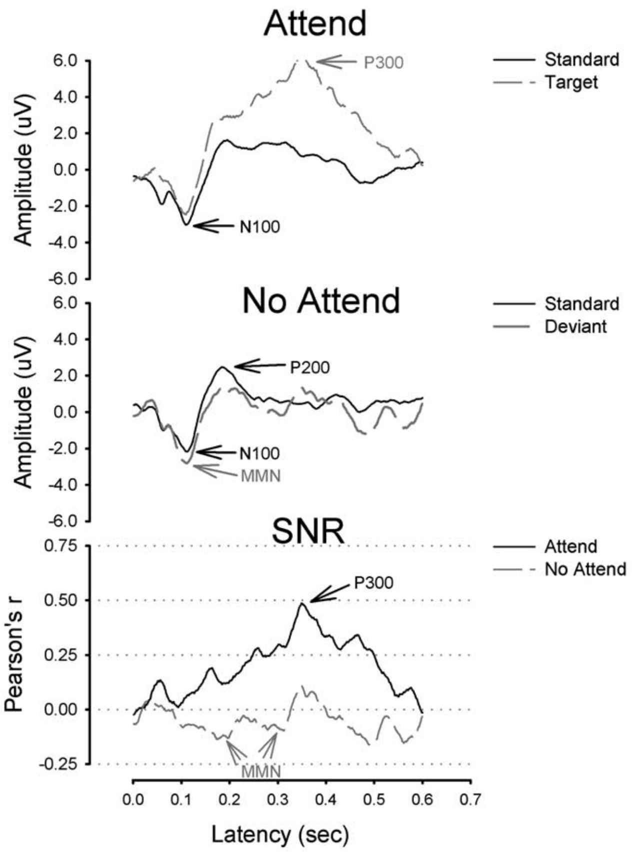

Fig. 1. Example of auditory ERP waveforms from an individual participant at a central electrode locations (Fz) during attend condition (top) and no attend conditions (middle). Signal-to-noise ratio (bottom) plotted as Pearsons's r on a point-by-point basis over time is also shown for attend and no attend conditions.

Table 1
Selected Components of the Auditory ERP[a]

Participants	P300	MMN	N100
BB	0.48	−0.13	0.28
DE	0.38	Absent	0.37
DM	0.06	−0.19	0.43
JC	0.60	−0.02	0.34
LZ	0.32	−0.03	0.38
LC	0.24	−0.14	0.25
LZ	0.49	−0.14	0.33
ME	0.11	−0.21	0.11
RB	0.22	−0.13	0.17
TV	0.18	−0.19	0.31

[a] Pearson's r-values for individual participants. Signal-to-noise ratio for N100 was computed as the maximum r-value from 80 to 160 ms; SNR for P300 was computed as the maximum negative r-value from 250 to 500 ms; SNR for MMN was computed as the maximum positive r-value from 100 to 200 ms.

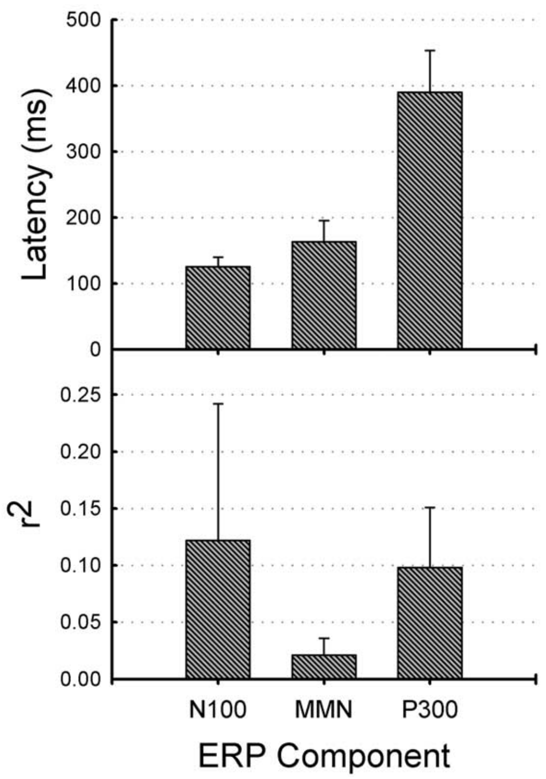

Fig. 2. (Top) Bar graphs (mean + 1 SD) for latencies of individual ERP components; Bottom) Bar graphs (mean + 1 SD) for r^2 of the individual ERP components. Significant differences in SNR were found between ERP components based on ANOVA ($F = 4.97$, $P < 0.019$).

and 2; Table 1). Low SNR of MMN is often acknowledged in the literature but typically not measured (e.g. Refs. [13, 17]). As an example, Pekkonen et al. [17] compared MMN and N100, but this comparison did not involve actual measurement of SNR. Nevertheless, it was concluded that MMN and N100 could be used both at the group and individual level. However results from the present study, using an explicit measure of SNR, suggests otherwise. With respect to MMN, McGee et al. [13] (p. 653), put the issue of SNR in perspective, "The SNR is sufficient for obtaining group averages, but responses from individual subjects are difficult to interpret. Thus, the described procedures result in responses that may not be useful for studies in which measurements from individuals are at issue". Dalebout and Fox [4] echo this viewpoint as their data demonstrate that MMN was not consistently present within individuals. If MMN measured within individuals cannot be assessed on a consistent basis, then test re-test reliability will be low, the validity of the metric will be questionable and potential clinical applications will be limited. Kurtzberg et al. [10] studied SNR of MMN whereby an F ratio was computed at a single point (F_{sp}) in the AEP waveform [5]. These authors found that F_{sp} values were highly variable but qualified their assessment by noting that the F_{sp} method was less applicable to cortical ERPs compared with ABRs due to relative small numbers of trials being analyzed. Because important clinical applications might be derived from MMN, having a straightforward measure of SNR would provide a more objective means to study MMN and therefore would facilitate the search for optimizing stimulus and recording parameters.

Signal-to-noise ratio is simply the magnitude of the signal divided by the magnitude of the noise. When computed over trials, this can be quantified as the mean of the signal divided by the standard deviation of the signal [18]. However the mean value might not always be the best characterization of the signal, as would be the case with an amplitude modulated carrier frequency. In the case of the MMN or P300, the response is often defined as the difference between two conditions; the standard vs. the deviant or the standard vs. the target. Likewise, the N100 response can be conceptualized as the difference between a baseline condition and a stimulus condition. Thus, a statistic dealing with a mean difference is particularly appropriate for these measures. As noted above, Kurtzberg et al. [10] assessed SNR by means of the F ratio. In contrast, the present study used Pearson's r. Given two alternatives (standard vs. deviant or standard vs. target) several statistical indices can be computed which are essentially interchangeable. For example, a value of t can be computed from r:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

which in turn can be squared to produce F. Likewise, F can be computed from r^2. In terms of experimental design considerations, the values of F and r^2 have the advantage of applying to cases involving three or more alternatives. Furthermore, since r^2 is the percentage of variance of y accounted for by x, it is more tractable statistically. Moreover, the values of t and r have the advantage of retaining the sign of the difference. Thus, in the present study, MMN was initially defined as the maximum negative value of r in the 100–200 ms interval. Had we used r^2, large positive deflections would also have been included. For statistical comparisons of r-values, Fisher's z transformation can be applied. As an alternative, we used the coefficient-of-determination, r^2. This transformation allowed for direct statistical comparisons with P300, since at this juncture, we were only concerned with the magnitude of the effect. Nevertheless, these several measures mentioned above can be used somewhat interchangeably to assess the normalized magnitude of the ERP.

Increasing the number of target and deviant stimulus trials is one pragmatic method often cited for improving SNR (i.e., to decrease the noise within n recording epochs). This relatively simple methodological consideration is fundamental to basic models of signal averaging in the time domain which assume that data contained within individual trials are composed of a linear combination of time invariant phase-locked responses plus background noise. However, as applied to MMN, this model is constrained in large part by the assumption of stationarity. Consequently, if MMN is not precisely phase-locked to the stimulus and/or if the response varies over time, then signal averaging in the time domain may not be the most appropriate method-of-analysis. Time domain-processing techniques, such as linear averaging, remove non-phase-locked EEG activity from the composite average by phase cancellation [12,16]. Consideration of this process may be significant because recent experimentation using both passive and active oddball paradigms suggests there is a substantial portion of event related energy retained in the response that is not phase-locked to the stimulus [3]. Moreover, there is evidence suggesting that habituation of MMN occurs over time [13]. If this is true, then simply averaging more responses could actually be counter productive.

The present results suggest that SNR of MMN needs to be improved if this ERP is to be used as a measure of auditory processing in individuals. Improvements could involve different signal processing strategies, which might include spectral analysis [2]. Alternatively, it may be necessary to develop better recording procedures. In either case, the use of a normalized statistic such as Pearson's r will facilitate the evaluation of alternative methodologies.

Stimulus conditions in which easily discriminable differences exist between standard and deviant stimuli show that MMN is not as robust as other averaged long latency auditory ERPs. Pearson's r, or the alternative r^2, provides a useful index of SNR in ERP waveforms. This approach to studying SNR is one method to evaluate alternative ways of improving SNR of MMN in individuals.

References

[1] A.T. Cacace, D.J. McFarland, Quantifying signal-to-noise ratio of the mismatch negativity, Assoc. Res. Otolaryngol. 25 (2002) 145.

[2] A.T. Cacace, D.J. McFarland, Spectral dynamics of electroencephalographic activity during auditory information processing, Hear. Res. 176 (2003) 112–132.

[3] A.T. Cacace, D.J. McFarland, Characterizing phase-locked versus non-phase-locked EEG activity in passive and active oddball paradigms, Assoc. Res. Otolaryngol. 26 (2003) 228.

[4] S.D. Dalebout, L.G. Fox, Reliability of mismatch negativity in individual listeners, J. Amer. Acad. Audiol. 12 (2001) 245–253.

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

[6] C. Escera, C. Grau, Short-term replicability of mismatch negativity, Electroenceph. Clin. Neurophysiol. 100 (1996) 549–554.

[7] H. Jasper, The ten-twenty electrode system of the International Federation, Electroenceph. Clin. Neurophysiol. 10 (1958) 371–375.

[8] T. Kujala, R. Näätänen, The mismatch negativity in evaluating central auditory dysfunction in dyslexia, Neurosci. Biobehav. Rev. 25 (2001) 535–543.

[9] N. Kraus, T. McGee, A. Micco, T. Carrell, A. Sharma, T. Nicol, Mismatch negativity in school-age children to speech stimuli that are just perceptually different, Electroenceph. Clin. Neurophysiol. 88 (1993) 123–130.

[10] D. Kurtzberg, H.G. Vaughan Jr, J.A. Kreuzer, K.Z. Fliegler, Developmental studies and clinical application of mismatch negativity: Problems and prospects, Ear Hear. 16 (1995) 105–117.

[11] H. Lang, O. Eerola, P. Korpilahti, I. Holopainen, S. Salo, O. Aaltonen, Practical issues in the clinical application of mismatch negativity, Ear Hear. 16 (1995) 117–129.

[12] S. Makeig, Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones, Electroenceph. Clin. Neurophysiol. 86 (1993) 283–293.

[13] T.J. McGee, C. King, K. Tremblay, T.G. Nicol, J. Cunningham, N. Kraus, Long-term habituation of the speech-elicited mismatch negativity, Psychophysiol. 38 (2001) 653–658.

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

[15] R. Näätänen, Attention and Brain Function, Lawrence Ehrlbaum Associates, New Jersey, 1992.

[16] C. Pantev, Evoked and induced gamma-band activity of the human cortex, Brain Topogr. 7 (1995) 321–330.

[17] E. Pekkonen, T. Rinne, R. Näätänen, Variability and replicability of the mismatch negativity, Electroenceph. Clin. Neurophysiol. 96 (1995) 546–554.

[18] S.W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, California Technical Publishing, San Diego, 1998.