J Am Acad Audiol 7: 105–112 (1996)

Using Naturally Produced Speech to Elicit the Mismatch Negativity

Sharon A. Sandridge*
Arthur Boothroyd†

Abstract

The mismatch negativity (MMN) was recorded from 10 young adults with normal hearing using naturally produced speech contrasts. Consonant-place and vowel-height contrasts were examined in consonant-vowel (CV) syllables by pairing either the consonant /t/ or /p/ with the vowel /I/ or /E/. Vowel-height was also examined as a pseudovowel; one cycle of the vowel segment of a CV was extracted and replicated over 200 msec. A total of five contrasts were examined: vowel-height following /p/, vowel-height following /t/, consonant-place preceding /E/, consonant-place preceding /I/, and pseudovowels. Significant MMN responses were elicited from all five contrasts, albeit with different amplitudes, latencies, and waveform morphology. The pseudovowel elicited the most well-defined MMN with the greatest amplitude and was found to be significantly different from the other four contrasts. Naturally produced speech stimuli appear to be viable stimuli for eliciting the MMN.

Key Words: Event-related potentials, mismatch negativity (MMN), speech

The mismatch negativity (MMN) is a response to a physically deviant stimulus in a train of homogeneous stimuli occurring approximately 100 to 300 msec post-stimulus onset in the auditory event-related potential (Näätänen et al, 1978). The response is elicited in the absence of attention (Näätänen, 1990) and reflects a physical mismatch between a memory trace formed by the standard stimuli and the incoming features of the deviant stimuli (Näätänen et al, 1978). The MMN has been elicited by changes in such parameters as frequency (e.g., Sams et al, 1985), duration (e.g., Kaukoranta et al, 1989), intensity (e.g., Näätänen et al, 1987), and location (e.g., Paavilainen et al, 1989). Besides simple tonal stimuli, MMN can be elicited with more complex stimuli such as complex tonal sequences (e.g., Schröger et al, 1992) or speech (e.g., Aaltonen et al, 1987). It can even be elicited when the differences between stimuli are near psychophysical thresholds (Sams et al, 1985), although the degree of physical difference between the stimuli affects the MMN response. As the difference increases, the MMN peak latency and the duration shorten, the magnitude of the amplitude increases (Näätänen and Gaillard, 1983) and the morphology changes (Scherg et al, 1989).

Because the MMN is an automatic response (i.e., requires no attention to the task) and reflects a neurophysiologic index to fine acoustic differences, it has potential as an objective tool for the assessment of speech perception, especially in clinical populations where standard speech perception measurements are contraindicated. For example, Kraus et al (1993a) investigated the clinical applicability of the MMN in two clinical populations. When synthesized speech pairs of /ta-da/ were presented to subjects with cochlear implants, responses from "good" cochlear implant patients were "strikingly similar to those recorded from normal listeners" (Kraus et al, 1993a, p. 123), suggesting the feasibility of MMN in a hearing-impaired population.

In another report, variants of synthesized pairs /ga-da/ and /ba-wa/, rated as "easy, hard, and hardest" for discrimination, served as the stimuli for MMN testing as one test in a central auditory processing test battery. The subject had normal peripheral hearing sensitivity but reported complaints of hearing difficulty,

*Department of Otolaryngology and Communicative Disorders, Cleveland Clinic Foundation, Cleveland, Ohio; †Graduate School, City University of New York, New York, New York

Reprint requests: Sharon A. Sandridge, Department of Otolaryngology and Communicative Disorders, The Cleveland Clinic Foundation, A71/9500 Euclid Avenue, Cleveland, OH 44195

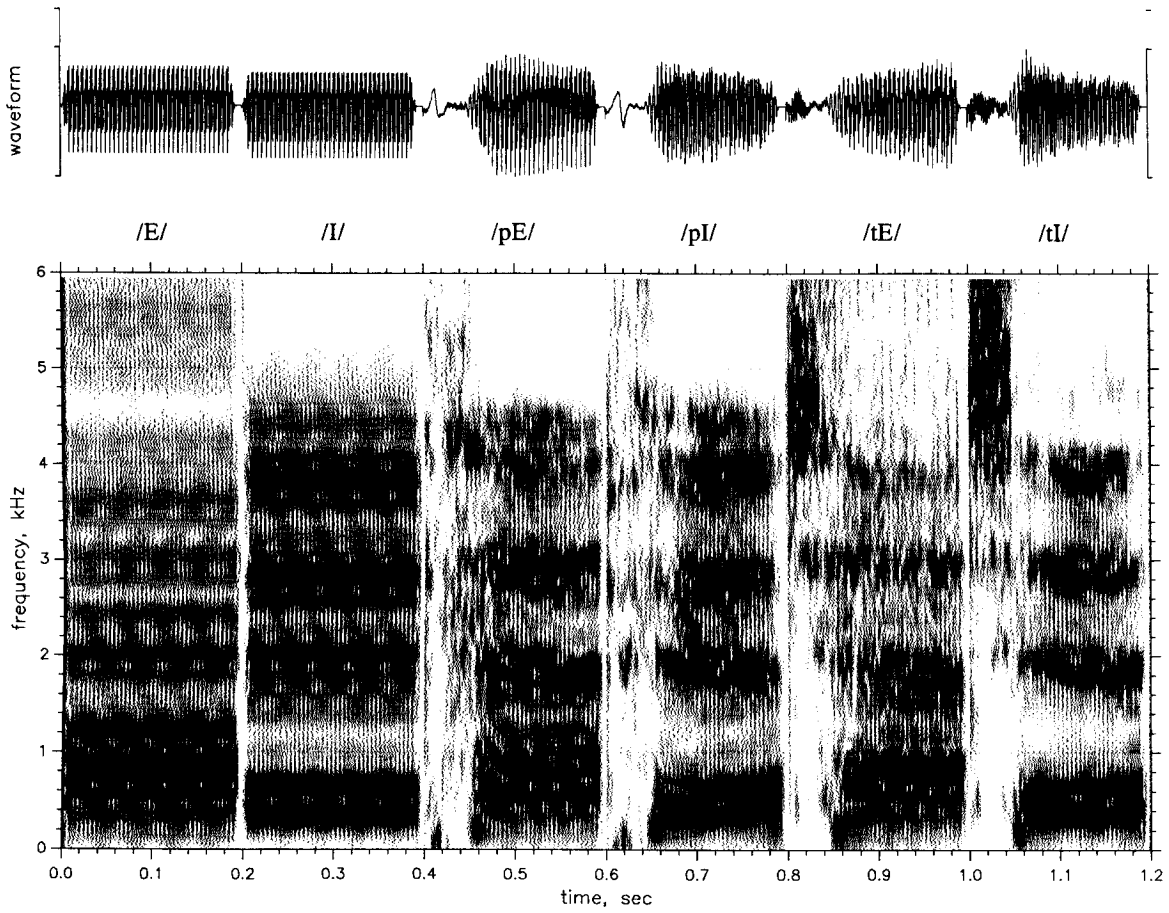

Figure 1 Spectrograms of the six stimuli used to examine the five speech contrasts. The top display illustrates the waveform while the bottom is the spectrographic display.

especially in a background of noise. When the pair contrasts were discriminated behaviorally, MMNs were recorded. When the pairs fell below the chance level behaviorally, MMNs were not elicited (Kraus et al, 1993b).

While the MMN shows promise as a clinical tool, further research needs to be conducted to establish a protocol that is clinically feasible. One such issue that needs to be investigated is the use of natural speech as the stimulus rather than synthesized speech, the choice by many investigators (Aaltonen et al, 1987, 1992; Sams et al, 1990; Kraus et al, 1992, 1993a, b, c, d; Sharma et al, 1993). While the use of synthetic speech offers obvious advantages over natural speech (e.g., control of acoustic differences), the use of natural speech may be a more ecologically significant stimulus. While intuitively the use of natural speech should elicit an MMN (for the basic reason that MMNs can be elicited with any acoustic difference), the effect of natural speech with its naturally occurring variations on the MMN has yet to be investigated. Likewise, the typical speech pairs previously used have differed in consonant-place using voiced stop consonants or voicing using alveolar stop consonants. A few studies have varied the vowels but none of the studies have systematically compared several speech contrasts within the same study. Thus, in addition to using naturally produced speech as the stimuli, three naturally occurring speech contrasts paired with two different vowels or two different consonants will be investigated.

SUBJECTS

Ten young adults (two males and eight females) ranging in age from 25 to 40 years served as subjects. All subjects had hearing sensitivity within normal limits and demonstrated the ability to correctly identify the speech contrast of interest with 100 percent accuracy.

STIMULI

The stimuli were naturally produced by a female talker, recorded and digitized at 16 bits and 30 kHz. Three speech contrasts were examined using two stop consonants (/t/ and /p/) and two vowels (/E/ and /I/) presented in either a consonant-vowel format (CV) or as isolated vowels. Vowel-height in natural CV syllables was examined by comparing /pE-pI/ and /tE-tI/ and their reversed pairs, /pI-pE/ and /tI-tE/. Consonant-place in natural CV syllables was examined using /pE-tE/ and /pI-tI/ and their reversed pairs, /tI-pI/ and /tE-pE/. The third speech contrast examined natural vowel-height in isolation when using pseudovowels. To produce these stimuli, one cycle from the vowel /I/, in CV /tI/, was extracted and replicated over 200 msec, yielding a complex, steady-state tone with a natural vowel spectrum. The procedure was repeated for the vowel /E/ in the CV /pE/. Speech spectrograms are shown in Figure 1 for the six stimuli.

The CV syllables were truncated to 200 msec and voice onset times were equated at 70 msec. The pseudovowels were matched for fundamental frequency and rms amplitude. Other naturally occurring variations were allowed to occur and were not controlled.

Each stimulus served as a "standard" (frequently occurring, indicated as the first stimulus in the pair) for one combination and also as a "deviant" (infrequently occurring, noted as the second stimulus in the pair) for the reversed contrast combination. For example, in the condition /pE-tE/, the /pE/ served as the standard and /tE/ served as the deviant stimulus. When the condition was reversed, in /tE-pE/, the standard was now the /tE/ while the deviant was the /pE/. See Table 1 for the contrast combinations.

Table 1 Stimuli Contrasts and Pairs

Natural Vowel-Height	
/p/V:	pE-pI
	pI-pE
/t/V:	tI-tE
	tE-tI
Consonant-Place	
C/E/:	pE-tE
	tE-pE
C/I/:	pE-tE
	tE-pE
Pseudovowel Height	
PV:	E-I
	I-E

The first stimulus in the pair served as the standard in that combination. The second stimulus served as the deviant stimulus.

PROCEDURES

The stimuli were presented at 80 dB SPL binaurally through ear inserts using the classic "oddball" paradigm, with the deviant stimuli presented 15 percent of the time and the standard presented 85 percent of the time. For each run, a total of 300 stimuli were presented (45 deviants and 255 standards), with an onset to onset (ISI) rate of 900 to 950 msec. The order of stimulus presentation within the run was pseudorandom, that is, the order was manipulated so that a minimum of three standards preceded every deviant and a deviant stimulus did not occur within the first 10 stimuli of the trial. Each condition was replicated three times (two subjects' responses were only replicated twice), yielding a possible total of 135 deviants.

Because each stimulus served as the deviant in one pair combination, and as the standard in another pair combination, two conditions were presented for each contrast, yielding a total of 10 conditions for five contrasts: vowel-height following /t/, vowel-height following /p/, consonant-place preceding /E/, consonant-place preceding /I/, and pseudovowels. The presentation order of the combinations was random across subjects.

Subjects were tested in an acoustically and electrically treated room while seated in a recliner reading a book of their choice. They were encouraged to ignore what they heard and to pay close attention to their reading material. Short breaks were provided between each condition with a longer break provided at 1-hour intervals. Total testing time was approximately 3 hours.

ELECTROPHYSIOLOGIC RECORDING AND ANALYSIS

Gold-cup electrodes were placed at Fz, Cz, Pz, A1, and A2 and referenced to the tip of the nose. The ground was Fpz. Vertical eye movements were recorded between two electrodes placed above and below the subject's right eye. Electrode impedance was less than 5000 ohms for the entire test session.

The EEG activity was amplified, filtered (0.1–100 Hz), collected over an 853-msec time epoch (including a 50-msec prestimulus period), and stored for offline processing. Offline activity

involved baseline correction, artifact rejection (±100 μV), digital filtering (0.1–30 Hz), and averaging of the EEG.

Separate averages were computed for the standard and deviant stimuli. Standard stimuli that immediately followed a deviant stimulus, as well as the first five standard stimuli in the block, were not included in the averaging of the standard stimuli.

Replicated responses were averaged together yielding an individual average waveform for each condition. The standard waveforms for each pairing of a contrast (i.e., /pE-tE/ and /tE-pE/) were then averaged, yielding a total of five waveforms for each individual for the standard stimuli and five waveforms for the deviant stimulus (see Table 1). It was assumed that, by averaging the combination and its reversal, naturally occurring variations other than those responsible for phonetic contrasts would be averaged out. The individual waveforms were then averaged to derive group waveforms. From the group waveforms, difference waveforms were computed by subtracting the standard grand individual average waveform from the deviant grand individual average waveform for each contrast.

The MMN was identified from the difference waveforms based on the following criteria: a negativity seen within the time window of 100 to 300 msec, the negativity more pronounced in Fz and Cz than Pz, and an inversion of that response at the mastoid (A1 and A2). Figure 2 illustrates these criteria; note the inversion of the response for A1 and A2 electrode sites, with the response in Pz showing reduced negativity.

DATA ANALYSIS

Statistical testing was performed to assess (1) if the standard and deviant waveforms differed from each other and across the conditions; (2) if the MMN amplitude in the 100-msec window around the peak was different from zero; and (3) if the MMN was different across the speech contrasts. Because it was not the purpose of this project to compare responses at different electrode sites, only Fz was submitted to statistical testing. Fz was used since the MMN is maximally recorded from the frontocentral electrodes (Näätänen et al, 1982; Sams et al, 1985).

For statistical testing, the waveforms were divided into segments representing smaller time windows with each segment averaged. To test differences between standard and deviant waveforms, the time window between 100 to 300 msec was divided into 20 segments. For the difference waveforms, the 100-msec time window (±50 msec on either side of the most negative peak within the 100- to 300-msec time window) was divided into 10 segments. The average of each segment from the standard and deviant waveforms was then submitted to an analysis of variance for repeated measures (ANOVA-R), with follow-up post hoc analyses using the Tukey Honestly Significantly Different tests. T-tests for dependent means were performed on the average of each segment from the difference waveform.

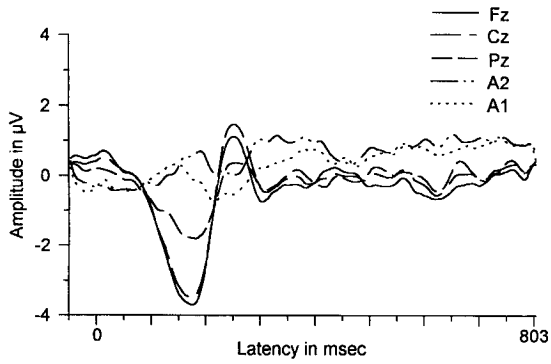

Figure 2 Grand mean difference waveform for the pseudovowel contrast illustrating the criteria used to identify the MMN response.

RESULTS

Responses from the group data are shown in Figure 3. The grand mean waveforms for the standard and deviant stimuli for the electrode site Fz for the five contrasts are shown along with the corresponding difference waveforms.

Standard versus Deviant Waveforms

The standard and deviant stimuli elicited classic N_1 and P_2 deflections followed by a flattening of the response for the remainder of the time window (Table 2). The latency of the N_1 deflection was similar between the standard and deviant waveforms. The P_2 deflection to the deviant stimuli had a similar peak latency as the standard stimuli but with a reduced amplitude. The extent of the amplitude differences varied across conditions (see Fig. 3). The amplitude measurements were baseline-to-peak for the N_1 and P_2 deflections and are shown in Table 2.

Using the average of the 20 segments within the 100- to 300-msec time window, a two-way ANOVA for repeated measures (waveform

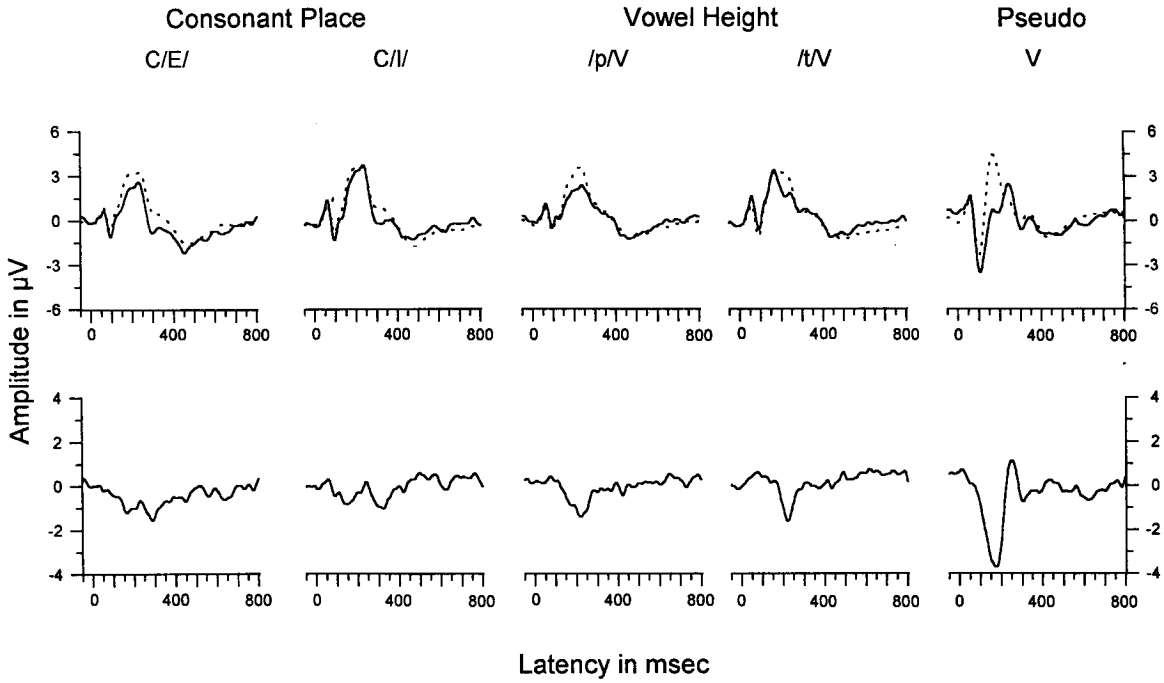

Figure 3 The top five graphs illustrate the standard (dotted line) waveforms and the deviant (solid line) waveforms from the group means for the five contrasts. The bottom five are the corresponding difference waveforms. Note the negative deflection occurring between 100 and 300 msec in the difference waveform.

[standard vs deviant] × contrast [vowel-height following /t/, vowel-height following /p/, consonant-place preceding /E/, consonant-place preceding /I/, and pseudovowels]), for Fz only, indicated main effects for waveform and contrasts as well as a significant interaction between the waveforms and contrasts (see Table 3 for ANOVA summary).

Post hoc testing for waveform indicated that the standard waveforms were significantly different from the deviant waveforms for three contrasts for the time window of interest: vowel-height following /p/ ($p < .05$), consonant-place preceding /E/ ($p < .000$), and pseudovowels ($p < .0001$). Vowel-height following /t/ and consonant-place preceding /I/ showed differences between the standard and deviant waveforms that were not less than the .05 significance level.

No significant differences across the contrasts for the standard waveforms were shown through post hoc testing for contrasts. There were, however, several significant differences across the contrasts for the deviant waveforms. The pseudovowel deviant waveform was significantly different from deviant waveforms of the other four contrasts (vowel-height following /p/

Table 2 Peak Latency and Amplitude* Measures of N_1 and P_2 for Standard and Deviant Waveforms for Each Condition

	N_1				P_2			
	Deviant		*Standard*		*Deviant*		*Standard*	
	LAT (msec)	*AMP (µV)*	*LAT (msec)*	*AMP (µV)*	*LAT (msec)*	*AMP (µV)*	*LAT (msec)*	*AMP (µV)*
C/E/	91	−1.15	91	−0.93	215	2.49	210	3.29
C/I/	91	−1.55	93	−1.01	222	3.52	203	3.62
/p/V	93	−0.52	93	−0.74	236	2.38	225	3.30
/t/V	92	−1.08	92	−1.26	168	3.20	190	3.40
PV	102	−3.5	98	−2.42	168	0.74	168	4.44

*Baseline-to-peak amplitude measurement.
LAT = latency; AMP = amplitude.

Table 3 Results of the Two-way ANOVA-R for Differences for Waveform

Effect	*df Effect*	*msec Effect*	*df Effect*	*msec Effect*	*F*	*p-level*
Waveform	1	29.8	38	6.47	4.6	.038
Contrast	4	6.9	152	.45	15.1	.000
Waveform X Contrast	4	1.6	152	.45	3.6	.007

$p < .001$, vowel-height following /t/ $p < .001$, consonant-place preceding /E/ $p < .003$, consonant-place preceding /I/ $p < .001$). The consonant-place preceding /E/ contrast was found to be significantly different from its counterpart, consonant-place preceding /I/ ($p < .003$). The vowel-height contrasts following either /t/ and /p/ were not significantly different from each other.

Difference Waveforms

The presence of the MMN was determined statistically by five separate t-tests for dependent means. The MMN amplitude was significantly different from the zero line for all contrasts (Table 4). The greatest amplitude (–2.77 μV) was obtained for the pseudovowel contrast. The vowel-height contrasts in CVs yielded amplitudes of –1.09 μV and –0.89 μV (following /p/ and /t/, respectively), while the amplitudes of the consonant-place contrasts were –0.89 μV and –0.59 μV (preceded by /E/ and /I/, respectively). Statistical significance was reached for the pseudovowel contrast only. It was different from the other four contrasts ($p < .001$ for all conditions).

DISCUSSION

The MMN was obtained for all naturally produced speech contrasts. The responses, however, were different among the speech contrasts. Both the pseudovowels and vowel-height in CV contrasts elicited well-defined MMNs characterized by a single negative peak. The pseudovowels, however, elicited the MMN with the greatest amplitude and most well-defined morphology. This finding is not unexpected for two reasons. First, the acoustic differences for the pseudovowels were present at onset and were maintained throughout the duration of the stimuli. The other contrasts, on the other hand, had naturally occurring changes of amplitude, spectrum, and fundamental frequencies throughout the time course. Second, there was more individual variability for the other contrasts compared to the pseudovowel contrasts, which resulted in reduced amplitude and a wider response. For example, the individual peak latencies fell within a range of 120 msec for the natural vowel contrasts compared to a range of only 40 msec for the pseudovowel contrast. This resulted in a widening of the deflection and a reduction of the amplitude of the grand mean waveform.

Table 4 Mean Amplitude (μV) along with Results from the t-tests for the Difference Waveforms for Each Condition

Condition	*Amplitude*	*t-test*	*df*	*p value*
C/E/	–0.89	10.8	9	.000
C/I/	–0.59	9.3	9	.000
/p/V	–1.09	15.2	9	.000
/t/V	–0.89	5.2	9	.001
PV	–2.77	8.6	9	.000

For the consonant-place contrasts, a different morphology pattern existed. Unlike the vowel contrasts, which yielded well-defined single deflections, the MMN for the consonant-place contrasts was characterized by a bifurcated negative deflection. The negative deflection began around 70 msec, peaking around 150 msec, followed by a positive-going slope (not crossing the baseline) peaking around 230 msec, and then falling to a second negative peak around 300 msec (see Fig. 3). This double peak in the MMN response was consistent for both consonant-place contrasts and may reflect a response to the two components of this contrast, namely, spectral differences in the burst and spectral differences in the vowel onset. Alternately, the double peak may be reflective of the two-component MMN model postulated by Scherg et al (1989). Regardless, the possibility that electrophysiologic methods can demonstrate separate responses to different contrast correlates of complex phonetic contrasts is exciting and warrants further research.

Table 5 Peak Latency and Amplitude Measures from the Difference Waveforms for Each Condition

	*Peak 1**		*Peak 2*		*Midpoint*	
Condition	*Latency (msec)*	*Amplitude (μV)*	*Latency (msec)*	*Amplitude (μV)*	*Latency (msec)*	*Amplitude (μV)*
C/E/	161.7	−1.33	286.7	−1.57	221.7	−1.37
C/I/	151.7	−1.01	315.0	−0.99	231.7	−0.90
/p/V	213.3	−1.44				
/t/V	218.3	−1.63				
PV	170.0	−3.80				

*Peak 1 latency values used for statistical testing.
Note that for the consonant-place contrasts there are two peaks identified.

To control for any naturally occurring variations that were not responsible for the phonetic contrasts, a stimulus and its pair were combined. For example, for the vowel-height following /p/ contrast (/pE/-/pI/), the responses to the /pE/ and /pI/, when they served as standards, were averaged together. Likewise, the same procedure was followed when they served as deviant stimuli. While this procedure may have averaged out the unwanted variations, it is possible that the responses may have been degraded. That is, by combining the responses to /pE/ as a deviant with /pI/ as a deviant, the inherent differences may have been lost and the combined response may have reflected the "jitter." A cursory review of the individual contrasts did not reveal an obvious difference between the pairs within the contrasts; however, this should be investigated further if this procedure is used for future research.

In conclusion, the MMN was reflective of the acoustic differences that occurred naturally in the CVs — differences that were not controlled. Although the MMN was present for natural syllables, it was weaker and less well defined than for pseudovowels. The pseudovowel may prove to be a viable stimulus in future MMN research investigating electrophysiologic assessment of speech perception and needs further investigation. However, if this test is to be used clinically, it will be necessary to demonstrate significant responses in individual subjects. Future research should concentrate on verifying the response in individual data.

Acknowledgment. Portions of this paper were presented at the 1994 American Academy of Audiology Convention, Richmond, VA.

This project was supported by grant no. DC000178 from the National Institutes of Deafness and Other Communication Disorders.

REFERENCES

Aaltonen O, Niemi P, Nyrke T, Tuhkanen M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biol Psychol* 24:197–207.

Aaltonen O, Paavilainen P, Sams M, Näätänen R. (1992). Event-related brain potentials and discrimination of steady-state vowels within and between phoneme categories: a preliminary study. *Scand J Log Phon* 17:107–112.

Kaukoranta E, Sams M, Hari R, Hämäläinen M, Näätänen R. (1989). Reactions of human auditory cortex to a change in tone duration. *Hear Res* 41:15–21.

Kraus N, Micco AG, Koch DB, McGee T, Carrell T, Sharma A, Wiet RJ, Weingarten CZ. (1993a). The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. *Hear Res* 65:118–124.

Kraus N, McGee T, Carrell T, Sharma A, Micco A, Nicol T. (1993c). Speech-evoked cortical potentials in children. *J Am Acad Audiol* 4:238–248.

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

Kraus N, McGee T, Micco A, Sharma A, Carrell T, Nicol T. (1993d). Mismatch negativity in school-age children to speech stimuli that are just perceptibly different. *Electroencephalogr Clin Neurophysiol* 88:123–130.

Kraus N, McGee T, Sharma A, Carrell T, Nicol T. (1992). Mismatch negativity event-related potential elicited by speech stimuli. *Ear Hear* 13:158–164.

Näätänen R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav Brain Sci* 13:201–288.

Näätänen R, Gaillard AWK. (1983). The orienting reflex and the N_2 deflection of the event-related potential (ERP). In: Gaillard AWK, Ritter W, eds. *Tutorials in ERP*

Research: Endogeneous Components. Amsterdam: Elsevier, 119–141.

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

Näätänen R, Paavilainen P, Alho K, Reinikainen K, Sams M. (1987). The mismatch negativity to intensity changes in an auditory stimulus sequence. In: Johnson R, Rohrbaugh RW, Parasuraman R, eds. *Current Trends in Event-Related Potential Research. Electroencephalogr Clin Neurophysiol (Suppl)* 40:129–130.

Näätänen R, Simpson M, Loveless NE. (1982). Stimulus deviance and evoked potentials. *Biol Psychol* 14:53–98.

Paavilainen P, Karlsson ML, Reinikainen K, Näätänen R. (1989). Mismatch negativity to change in spatial location of an auditory stimulus. *Electroencephalogr Clin Neurophysiol* 73:129–141.

Sams M, Aulanko R, Aaltonen O, Näätänen R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *J Cog Neurosci* 2:344–357.

Sams M, Paavilainen P, Alho K, Näätänen R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol* 62: 437–448.

Scherg M, Vajsar J, Picton TW. (1989). A source analysis of the late human auditory potentials. *J Cog Neurosci* 1:336–355.

Schröger E, Näätänen R, Paavilainen P. (1992). Event-related potentials reveal how nonattended complex sound patterns are represented by the human brain. *Neurosci Lett* 146:183–186.

Sharma A, Kraus N, McGee T, Carrell T, Nicol T. (1993). Acoustic vs. phonetic representation of speech stimuli as reflected by the mismatch negativity event-related potential. *Electroencephalogr Clin Neurophysiol* 88:64–71.