Phonemes, intensity and attention: Differential effects on the Mismatch Negativity (MMN)

Michael D. Szymanski*, E. William Yund, and David. L. Woods

Department of Neurology (127)
VA Medical Center
Northern California Health Care System
150 Muir Rd.
Martinez, CA 94553 USA

Running Title: MMN for phonemes and intensity

*Present address: Department of Radiology
Biomagnetic Imaging Laboratory
UC San Francisco
513 Parnassus Avenue S-362
San Francisco, CA 94143-0628
mszyman@mail.radiology.ucsf.edu

Corresponding Author: E. William Yund
VA NCHCS
150 Muir Road
Martinez, CA 94553
(925)372-2296
yund@ebire.org

PACS: 43.64.Ri, 43.64.Sj, 43.64.Qh, 43.71.Pc
Abstract

Auditory event-related potentials (ERPs) to speech sounds were recorded in a demanding selective attention task to measure how the mismatch negativity (MMN) was affected by attention, deviant feature, and task relevance, i.e., whether the feature was target- or nontarget-type. With vowel-consonant-vowel (VCV) disyllables randomly presented to the right and left ears, subjects attended to the VCVs in one ear. In different conditions, the subjects responded to either intensity or phoneme deviance in the consonant. The position of the deviance within the VCV also varied, being in the 1st (VC), 2nd (CV), or Both (VC and CV) formant-transition regions. MMN amplitudes were larger for deviants in the attended ear. Task relevance affected the MMNs to intensity and phoneme deviants differently. Target-type intensity deviants yielded larger MMNs than nontarget-types. For phoneme deviants there was no main effect of task relevance, but there was a critical interaction with deviance position. The Both position gave the largest MMN amplitudes for target-type phoneme deviants, as it did for target- and nontarget-type intensity deviants. The MMN for nontarget-type phoneme deviants, however, showed an inverse pattern such that the MMN for the Both position had the smallest amplitude despite its greater spectro-temporal deviance and its greater detectability when it was the target. These data indicate that the MMN reflects differences in phonetic structure as well as differences in acoustic spectral-energy structure of the deviant stimuli. Furthermore, the task relevance effects demonstrate that top-down controls not only affect the amplitude of the MMN, but can reverse the pattern of MMN amplitudes among different stimuli.

INTRODUCTION

The mismatch negativity component (MMN) of the event-related brain potential (ERP) has been interpreted as a response from a preattentive auditory change-detection system. The MMN was originally thought to reflect an acoustic spectral-energy mismatch between an incoming stimulus and the sensory memory trace of previous stimuli (Näätänen, Gaillard and Mäntysalo, 1978; Näätänen, 1990, 1992; Näätänen and Alho, 1995; Schröger, 1997). It is well established that MMNs can be generated in repetitive tone sequences by a change of one or more simple auditory features, including frequency, intensity, location, duration, and pattern (Näätänen, 1992).

Recent studies suggest that reliable MMNs are also generated by phonetic deviation. Näätänen, Lehtokoski, Lennes et al. (1997) and Dehaene-Lambertz (1997) demonstrated that the MMN is larger to natal language prototypical phonemes compared to foreign-language phonemes. Furthermore, Tremblay, Kraus, Carrell and McGee (1997) found that adults who were trained to discriminate foreign-language phonemes had larger MMNs to foreign-language deviant phonemes than did adults with no discrimination training. It is only the co-occurrence of unfamiliar pre-phonetic auditory features (or unfamiliar combinations of pre-phonetic auditory features) in foreign-language phonemes that mitigates against the stronger conclusion that the phoneme-specific MMN exists.

Näätänen and colleagues (Näätänen, 1992; Näätänen, Paavilainen, Tittinen et al., 1993; Näätänen and Winkler, 1998) have consistently maintained that attention is not required to elicit the MMN. Evidence supporting the automatic nature of the MMN includes the finding that the negativity occurs to deviant auditory stimuli even when a person’s attention is directed to another task in the auditory (Näätänen et al., 1978) or visual modality (Sams, Paavilainen, Alho and Näätänen, 1985; Woods, Alho and Algazi, 1992; Alho, Woods and Algazi, 1994). Typically, the MMN has been recorded while the subject was reading or watching a video recording. Although no
study to date has demonstrated a total absence of MMN in the absence of attention, numerous studies have shown that attended deviants evoke increased MMN amplitudes compared to ignored sounds. The specific features for which this attention effect has been found include intensity (Woldorff, Hackley and Hillyard, 1991), frequency (Näätänen et al., 1993; Oades and Dittmann-Balcar, 1995), duration (Alain and Woods, 1994), pattern (Alain and Woods, 1997; Nordby, Roth and Pfefferbaum, 1988a, 1988b), and tones embedded in a narrative (Trejo, Ryan-Jones and Kramer, 1995).

A confound in some earlier studies examining the effects of attention on the MMN (e.g. Woldorff et al., 1991) has been the possible inclusion of target-related deviant components (N2b) that may have contributed to the measured attentional modulation of MMN. To eliminate confounding N2b effects, some later studies have evaluated attention effects on nontarget deviants for stimulus features such as intensity (Näätänen et al., 1993), frequency (Trejo et al., 1995) and tonal pattern (Alain and Woods, 1997), and have nonetheless shown attention-related MMN enhancements. Attentional modulation of the MMN has been alternatively interpreted as the result of: 1) increased gain of the MMN generator for attended deviants (Näätänen et al., 1993), or 2) inhibition of MMN in the ignored channel (Trejo et al., 1995; Alain and Woods, 1997).

The effect of attention on the speech-evoked MMN has not been studied. There are reasons why the attentional modulation of the speech-evoked MMN might be expected to be either more or less than that of the non-speech MMN. Since speech is a complex high-information-content signal, attention might be expected to facilitate the phoneme MMN because more sophisticated information processing might be required. Conversely, since speech perception is highly overlearned and automatic, attentional modulation might be expected to be reduced (Kraus et al., 1996; Näätänen et al., 1997). Alternatively, if attention modulates the gain of the MMN generator in either an excitatory (Näätänen et al., 1993) or inhibitory (Trejo et al., 1995; Alain and Woods, 1997) manner, then the effect of attention should be essentially the same for speech and nonspeech signals.

There has been increased interest in using the MMN as a clinical diagnostic tool because this deviant-evoked negativity might provide an objective neural measure of auditory discriminability. One idea has been to employ the MMN as a test of the auditory system's ability to transmit the acoustic information important for understanding spoken language (Picton, 1995; Näätänen, 1995; Kraus, McGee, Carrell and Sharma, 1995). To date, the MMN has been used to assess the efficacy of phoneme discrimination training and related neural plasticity (Kraus, McGee, Carrell, King et al., 1995; Kraus, McGee, Carrell et al., 1996; Tremblay et al., 1997), and some investigators have proposed using MMN to monitor the effectiveness of hearing aid therapy (Picton, 1995) and cochlear implants (Ponton and Don, 1995).

The present study was designed to test the effectiveness of the MMN as an indicator of the transmission of the acoustic information needed for correct phoneme identification. The ultimate goal of the research was to develop MMN techniques for evaluating a signal-processing algorithm's ability to transmit auditory information through a sensorineural hearing impairment. The current study was designed to test the hypothesis that the MMN can be used to measure pre-phonetic acoustic information under different attention conditions and with variation in the behavioral task.

Evidence from previous MMN studies on phonetic information processing has confounded phonetic and pre-phonetic acoustic contributions to the MMN because the deviant stimuli have differed from the standards along corresponding acoustic and phonetic dimension. For instance, deviant vowels have different formant frequencies than standard vowels (Aaltonen et al., 1997; Dehaene-Lambertz, 1997; Näätänen et al., 1997), and deviant-consonant CVs have different formant transitions or voice-onset times than standard CVs (Kraus, McGee, Sharma et al., 1992; Sams,
Aulanko, Aaltonen and Näätänen, 1990; Sandridge and Boothroyd, 1996; Tremblay et al., 1997). In such studies, the observed MMNs could reflect either the acoustic or the phonetic deviation between vowels or consonants. The speech sounds used here were vowel-consonant-vowel (VCV) disyllables that included two different deviant features—phoneme and intensity—to reveal possible differences between MMNs to speech sounds elicited by phonetic and acoustic deviance. The middle consonant of the VCVs could differ in either phoneme (/b/ or /g/) or intensity (a more intense /b/). Each feature deviance also could occur in one of three temporal positions: during the 1st formant transition (VC), during the 2nd formant transition (CV), or during Both (VC and CV, combined). Comparing the MMNs to phoneme and intensity deviants for the three temporal positions might demonstrate differences between phonetic and acoustic MMNs that would not otherwise be apparent because of differences in the magnitudes of acoustic and phonetic deviance across the deviance positions. For phoneme and intensity deviants, the acoustic deviance would be greatest for the Both position. Thus, if phoneme and intensity MMNs were generated in response to simple acoustic deviation, then phoneme and intensity MMNs should be maximal with Both-position deviants and should be less for either 1st- or 2nd-position alone. Alternatively, if phoneme MMNs are generated by speech-specific phoneme deviance, then the Both-position phoneme-MMN would be expected to be smaller than either single-position MMN. This prediction is based on the fact that the Both-position phoneme deviant is less phonetically deviant than are the single-position phoneme deviants: In addition to the deviant consonant that is present in all phoneme deviants, the single-position phoneme deviants also include the conversion of the single middle consonant into a consonant pair—the stimulus is a VCV when the VC and CV transitions are phonetically consistent, and a VCCV when they are not.

I. METHODS

A. Subjects

The 13 subjects (five males and eight females) whose data were included in the analysis ranged in age from 20-39 years (mean 25.8). All were right-handed, except for one ambidextrous male and one left-handed female, and all had normal hearing (thresholds 20 dB HL or less from 250 to 4000 Hz). Subjects completed three experimental sessions, one behavioral session and two recording sessions. The purpose of the behavioral session was to train the subject to detect the various targets under simplified conditions and to verify that each subject could detect the targets under the demanding conditions to be used in the recording sessions (see Procedure, below). Subjects were paid for their participation and gave informed consent, according to DVA guidelines.

B. Stimuli

Vowel-consonant-vowel (VCV) stimuli were synthesized at a 20 kHz sampling rate with a Klatt speech synthesizer, KLSYN88a (MIT and Sensimetrics Corporation), in cascade mode with a 1 ms update interval. The digital stimulus waveforms were converted to analog signals (Data Translation DT2823), filtered (Tucker-Davis Technology, TDT, antialiasing filter FT6-92, 8 kHz), attenuated (TDT PA4) and delivered to Etymotic ER-2 earphones through a power amplifier (TDT HB6). The intensity of the steady-state portion of the vowels in the VCV stimuli was 75 dB SPL. The stimulus duration was 220 ms, including 10 ms linear onset and offset shaping.

The female-voice /ibi/, illustrated in Fig. 1, was the standard left-ear stimulus. Like this female-voice /ibi/, all left-ear stimuli had a fundamental frequency (F0) of 220 Hz and formant frequencies of 330 (F1), 2000 (F2), 3000 (F3) and 3600 Hz (F4) in the vowel. The right-ear male-voice /ibi/ differed only in F0 (156 Hz) and F1 (270 Hz). The difference in voice pitch was
included to assist the subjects in attending to stimuli in one ear and ignoring those in the other ear. Neither ear nor voice were intended to be a factor in the analysis, but separate analyses of left-ear-female-voice and right-ear-male-voice stimuli showed the same effects and interactions for the major factors of the analyses (attention, deviant feature, task relevance and deviance position—see below for factor definitions and descriptions).

There were two types of deviant features: (1) phoneme deviants, where /b/ was changed to /g/, and (2) intensity deviants, where the intensity of /b/ was increased (referred to as /B/). The end points for the formant transitions for /b/ were 1600 (F2), 2300 (F3) and 2800 (F4). Those for /g/ were 2395 (F2), 3000 (F3) and 3305 Hz (F4). In the intensity deviants, the intensity increases were restricted to the formant-transition regions of the VCVs, with the intensity (Klatt AV parameter) maxima occurring from 70-80 ms and at 130 ms. In a brief pilot experiment, a 15-dB increase in AV in these regions (phased in from 60-70 ms and out from 80-90 ms and 130-160 ms) was as detectable as the phoneme deviant. As already mentioned in the Introduction, the phoneme or intensity deviance could occur in three different temporal positions, (1) during the VC formant transitions, (2) during the CV formant transitions or (Both) during VC and CV formant transitions. Table 1 lists the different type of deviant stimuli and the temporal onsets of the deviance.

[Insert Table 1 about here.]

C. Procedure

The subject's task was to detect one type of deviant feature (intensity or phoneme) in one ear, independent of the position of the deviance. Since preliminary experiments indicated that it was difficult for inexperienced subjects to perform these detection tasks without some training, the experiment began with a training session to ensure that all subjects would be able to accurately detect the target deviants in the EEG recording sessions. In the initial training, stimuli were delivered to only one ear and at a slower rate than in the main experiment (200 ms added to the main-experiment inter-stimulus interval, ISI). The stimulus sequences used in training included the same proportion of intensity and phoneme deviants as in the main experiment. When the subject scored greater than 90% correct on the one-ear training for intensity and phoneme deviants, training proceeded to sequences with the same random right- and left-ear stimulus presentation and the same stimulus timing as in the main experiment. Subjects had to correctly detect at least 80% of each designated target with at most 5 false alarms (in a standard block of 800 total stimuli, 60 of which were targets), to continue on to the EEG recording sessions. This training was essential to be sure that subjects could perform the demanding experimental task.

EEG recording sessions were scheduled on two subsequent days. In each block of 800 stimuli subjects responded to one type of deviant stimulus in one ear (the target), while withholding responses to all other stimuli. Target detection was difficult and subjects spontaneously reported that they would miss targets if they failed to pay close attention to stimuli in the attended ear. Hand of response was counterbalanced across the two sessions for each subject, with subjects pressing the button on the top of a joystick with the thumb of the designated hand. Subjects were informed of their performance at the end of each block. The target was the same within each block and was varied across blocks in a pseudo-random order. Each subject completed 10 blocks for each target.

Before each block, the subject was informed of the target, and then the block began with 24 sample targets, eight for each deviance position. A three-second interval separated the sample targets from the test block of 800 random stimuli. During the test block, 400 stimuli were presented to each ear. The order was random with the constraint that each deviant type (right- or left-ear phoneme or intensity deviant) was followed by at least two other stimuli before it was repeated.
Stimuli in each ear consisted of 70% standards (/ibi/), 15% phoneme deviants (5% each of /igi/, /igbi/ and /ibgi/) and 15% intensity deviants (5% each of /iBbi/, /ibBi/ and /iBi/). The ISI in the test block was varied randomly between 80-125 ms in a rectangular distribution, such that the average rate of stimulus delivery was 3/s (Fig. 2).

D. EEG Recording

The EEG (bandpass 0.01 to 100 Hz) was recorded from 30 electrodes: BE (below left eye), LE (lateral to left eye), Nose, Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC6, LM (left mastoid), T7, C3, Cz, C4, T8, RM (right mastoid), CP5, CP6, P7, P3, Pz, P4, P8, PO5 (center of O1-P3-P7 triangle), PO6 (center of O2-P4-P8 triangle), O1 and O2. Electrode recordings were referenced to four interconnected, EKG-balanced electrodes at the base of the neck (Woods & Clayworth, 1985). The subject was grounded through an electrode at AFz.

The EEG was sampled at 256 Hz and continuously digitized to disk for later processing. The BE and LE recording channels were used to detect eye movements. During the off-line averaging, epochs with eye movements, signal clipping or excessive amplitude swings (greater than 75 μV peak-to-peak) were not included in the averages.

E. Data Analysis

Difference waves were obtained by subtracting ERPs for standard VCVs from those of deviants. The MMN amplitudes were measured relative to the 200 ms of prestimulus baseline EEG activity. The analysis of MMN scalp-distribution included 25 electrode sites, all except BE, LE, Nose, LM, and RM. The detailed analysis of MMN duration and amplitudes was done at the Fz electrode. In the analysis of MMN duration, mean amplitudes were measured in 20-ms time bins for each subject for each combination of the factors: attention (attended, ignored), deviant feature (phonetic, intensity), deviance position (1st, 2nd, Both) and task relevance (target-type feature, nontarget-type feature). With subject as the random variable, the voltage in each bin was tested (ANOVA) to determine if it was significantly different from zero. The duration of significant negativity (or positivity) was then defined as the time interval over which at least three consecutive 20-ms bins showed significance (at least two at p < 0.01).

Comparison across different factors and/or electrode sites requires an unbiased measure of amplitude. A source of difficulty was that the different deviant features and deviance positions produced MMNs with different latencies and somewhat different shapes. Therefore, to minimize bias in the MMN mean amplitude measurements, the temporal location of the 100-ms measurement interval was determined for each combination of deviant feature and deviance position, based on the intervals of significant negativity, determined as described above. Relative to stimulus onset, the measurement intervals for the intensity feature were 180-280 ms for Both and 1st positions and 240-340 ms for 2nd position. The measurement intervals for the phoneme feature were 200-300 ms (Both), 220-320 ms (1st) and 280-380 ms (2nd). These intervals were used for all MMN amplitude measurements for all 25 electrodes in the scalp-distribution analysis and for Fz in the detailed MMN amplitude analysis.

ANOVA for MMN amplitudes at the Fz electrode included subjects as the random factor and attention (attended, ignored), deviant feature (phonetic, intensity), deviance position (1st, 2nd, Both) and task relevance (target-type, nontarget type) as within-subjects factors. Separate ANOVAs were performed for the phonetic and intensity features because differences between these deviant features were expected, in particular for deviance position. Scalp distributions were evaluated across the 25 scalp sites, including Greenhouse-Geisser corrections assuming that the 25 sites represented
only 4 independent measures. Although this correction was used to determine significance levels, both the original degrees of freedom and the corrected degrees of freedom will be reported here for each analysis.

II RESULTS

A. Behavioral reaction times and accuracy

The overall mean reaction time (RT) was 397 ms with 87% of the targets detected and less than two false alarms (FAs) per block of 800 stimuli. Table 2 shows the RTs and hit rates broken down by deviant feature and deviance position. Significantly more deviant-phoneme targets were detected (90%) compared to deviant-intensity targets (83%) \[F(1,12)=10.56, p<0.01\]. The reverse trend for slower RTs for phoneme targets (402 versus 392 ms) was not significant. The temporal position of the deviance affected the hit rates and RTs of phoneme and intensity targets in similar ways. The Both position produced the highest hit rates \[F(1,12)=23.21, p<0.0001\] and the fastest RTs \[F(1,12)=219.78, p<0.0001\]. The performance difference between 1st- and 2nd-position targets was significant for hit rate \[F(1,12)=28.09, p<0.0001\], but not for RT.

B. ERPs for standard and deviant stimuli

ERPs for standard and deviant stimuli are illustrated in Fig. 3. Panels A and C include ERPs to stimuli in the attended ear, while B and D include ERPs to stimuli in the unattended ear. Similarly, A and B show ERPs when the attended intensity deviants were targets and C and D when the attended phoneme deviants were targets. Responses were averaged across ear of delivery and deviance position. As can be seen in Fig. 3, the ERPs to the standard VCVs were of low amplitude, never exceeding ±2.5 μV. This low-amplitude and relatively indistinct standard-VCV ERP is not surprising given the rapid presentation rate (3/s), the long VCV duration (220 ms) and the complex intensity and formant-frequency modulation that occurs in the VCV stimulus. This intensity modulation of the VCV was shown in Fig. 1: there is an onset at 0 ms, an offset at 90 ms, a second onset at 130 ms, and then the final offset at 220 ms; formant frequencies change in the time ranges 60-90 ms and 130-160 ms. In spite of these limited ERPs to the standard VCVs, clear negative and positive differences are present in the average ERPs to the deviant stimuli. The MMN was isolated in the difference waves by subtracting the ERPs for the standard VCV from the ERPs for the deviants. MMN onset occurred at 150-200 ms. This relatively long latency of onset reflects the fact that the standard and deviant stimuli were identical from 0-60 ms.

C. Time course of the negative-difference components

As expected, the time course of the MMN, as well as the positive response to detected targets (P3b), varied with the deviant feature and the deviance position. Table 3 gives the intervals over which the difference waves were significantly negative. Virtually all of the deviant stimuli generated broad negative differences in the MMN latency range. Although no further statistical tests were done to compare durations for different deviant stimuli or for different factors, some general patterns of time differences are apparent. When comparisons are made within the same deviant feature, Both- and 1st-position negativities tend to be similar and to start and/or end before 2nd-position negativities. Similarly, when corresponding deviance positions are compared, phoneme-deviant negativities tend to occur later than intensity-deviant negativities.
A strong attention effect was found for the MMN amplitudes measured at Fz in the 100 ms time intervals derived from the durations of significant negativity. The MMN had a greater mean amplitude for the attended ear compared to the ignored ear [F(1,12)=33.93, p<0.0005]. Stimuli in the ignored ear generated MMNs of about half the amplitude (-0.98 $\mu$V) of the MMNs of stimuli in the attended ear (-1.91 $\mu$V). The effect of attention on MMN amplitude is illustrated in Figs. 4 and 5, for nontarget phoneme and intensity deviants, respectively. These figures show the scalp distribution of the difference waves for attended and ignored nontarget-type deviants. The minimal-amplitude difference waves seen with the present reference-electrode system for the nose (Nz) and mastoid (RM, LM) electrodes indicates that these scalp distributions can be compared directly with those of other studies that used either nose or mastoid reference electrodes. Both deviant features yielded the fronto-central distribution typical of the MMN, with maximum amplitude (peak about 2 $\mu$V) at Fz, Cz, and adjacent sites (F3, F4, C3, C4, FC5 and FC6). Phoneme and intensity MMNs had similar amplitudes and showed similar enhancements with attention, consistent with the lack of a significant main effect of deviant feature and the lack of an interaction between the factors attention and deviant feature. Indeed, attention produced only a main effect, showing no significant interactions with any other factor—deviant feature [F(1,12)=1.48, n.s.], deviance position [F(1,12)=1.14, n.s.] or task relevance [F(1,12)<1.0].

[Insert Figure 5 about here.]

In the analysis of the scalp distribution of the MMN, the mean MMN amplitudes were normalized to equate amplitudes at Fz. This normalization procedure eliminates the main effects and interactions among the principal factors of the study (deviant feature, deviance position, attention, and task relevance), but permits an unbiased evaluation of the interactions of scalp location with these factors. As might be assumed from Figs. 4 and 5, the main effect of scalp location was highly significant for nontarget phoneme deviants [F(24,288)=10.74, p<0.001(3,36)] and for nontarget intensity deviants [F(24,288)=8.34, p<0.005(3,36)]. The interaction of attention and scalp location, however, was not significant for either phoneme [F(24,288)=1.36, n.s(3,36)] or intensity deviants [F(24,288)<1.0].

[Insert Figure 6 about here.]

E. Deviant feature, Deviance position and Task relevance

Since it was anticipated that the effect of deviance position might be different for the deviant features, phoneme and intensity, the remainder of the statistical analysis was carried out separately for these two features. Each of these analyses included the factors deviance position, task relevance and attention. Attention was included in these analyses to detect interactions with the other critical variables and, since there were none, it will not be discussed further. In the case of the phoneme feature, neither deviance position nor task relevance produced a significant main effect. However, their interaction was significant [F(2,24)=4.06, p<0.05] and is illustrated in Fig. 6. For the target-type phoneme deviant (cross-hatched bars), the Both position yielded a large amplitude MMN and the 1st position yielded the smallest amplitude MMN. In contrast, for the nontarget-type phoneme deviant (solid bars), the Both position yielded the smallest and the 1st position yielded the largest MMN. Post-hoc analyses (Tukey, Least Significant Difference Test) indicated that MMN amplitude for the 1st-position nontarget-type deviant was significantly greater than those for the 1st-position target-type and the Both-position nontarget-type deviants (each, p<0.05).

[Insert Figure 7 about here.]

The interaction between deviance position and task relevance for the phoneme deviants is also apparent in the difference waves from the Fz electrode, as shown in Fig. 7. The four panels of
the figure show the deviant-phoneme difference waves for: (A) targets, i.e. attended target-types; (B) attended nontarget-types; (C) ignored target-types; and (D) ignored nontarget-types. The larger MMNs in panels A and B compared to C and D represent the attention effect. The attention effect is unimportant in the present context, except to illustrate that the deviance-position-by-task-relevance interaction occurred independently of attention. The Both-position MMN has the largest amplitude in each of the target-type panels (A and C) and the 1st-position MMN has the largest amplitude in each of the nontarget-type panels (B and D). The occurrence of the interaction in the MMNs to ignored, as well as attended, deviants demonstrates that the source of the interaction is not in recognizing and responding to the target, but rather in the pre-response processing of target-type and nontarget-type deviants.

A different pattern of results was obtained for the intensity feature, as shown in Fig. 8. Here, there were significant main effects of task relevance \([F(1,12)=5.67, p<0.05]\) and deviance position \([F(2,24)=21.77, p<0.0005]\), but their interaction failed to reach significance \([F(2,24)=3.36, ns]\). In the effect of task relevance, target-type intensity deviants yielded larger MMN amplitudes than nontarget-types for each deviance position. For deviance position, the order of decreasing MMN amplitude was Both, 1st and 2nd. The similarity of 1st- and Both-position MMN amplitudes within target- and nontarget-type intensity deviants contrasts sharply with the crossover interaction found for the phoneme deviants and illustrated in Fig. 6.

The scalp-distribution analysis for deviance position (collapsed across attention and task relevance) also showed differences between the MMNs for intensity and phoneme features. For the intensity deviants, there was no significant interaction between electrode location and deviance position. For the phoneme feature, however, there was an interaction of electrode site by deviance position (SxP) \([F(48,576)=2.76, p<0.05(6,72)]\). The pattern of the SxP interaction for phoneme deviants is interesting because it was right-left asymmetrical. All 11 left-hemisphere electrodes showed the smallest MMN for the Both position, while 9 of 11 showed the largest MMN for the 1st position. MMN amplitude differences for the Both- and 1st-positions were less consistent in right-hemisphere electrodes, with 6 of 11 showing larger MMNs for the Both position. A secondary analysis was done to verify this right-left asymmetry in the deviance-position effect for the phoneme deviants, including the average MMN amplitudes for electrode sites F3, FC5, C3, CP5 and P3 representing the left hemisphere (LH), and F4, FC6, C4, CP6 and P4 representing the right hemisphere (RH). These electrode sites were chosen because they showed larger MMN amplitudes than other lateralized sites, as can be seen in Figs. 4 and 5. In the ANOVA, there was no significant main effect of deviance position \([F(2,24)<1.0]\) or...
laterality \[F(1,12)<1.0\], but the expected interaction was significant \[F(2,24)=5.76, p<0.01\] and is illustrated in Fig. 10. Post-hoc analyses (Tukey, Least Significant Difference Test) found significant differences for deviance position over the LH, with the Both-position MMN less than the 1st-position MMN \((p<0.001)\) and the 2nd-position MMN \((p<0.01)\). There were no significant differences among the RH MMNs. In comparisons between LH and RH response amplitudes, LH MMNs were significantly greater than RH MMNs for the 1st \((p<0.001)\) and the 2nd \((p<0.05)\) deviance positions, but the reversal (RH greater then LH) for the Both position failed to reach significance. The same secondary analysis was done for intensity deviants, but it will not be considered further because it showed only the main effect of deviance position already described above and illustrated in Figs. 8 and 9, with no laterality effect or interaction.

IV. DISCUSSION

There are three major findings in the present study: (1) the large, but relatively simple, non-interactive effect of attention on the MMN; (2) the non-interactive effects of the task relevance and deviance position on the MMN to intensity deviants; and (3) the complex interactions of task relevance and deviance position on the MMN to phoneme deviants. Although attention was not the primary focus of the present study, the results suggest a new explanation for the effects of attention on the MMN. An extensive discussion of attention effects seen in previous studies is required to evaluate this new hypothesis. The effects of task relevance and deviance position for the intensity deviants were quite simple and need to be discussed only briefly. These intensity-deviant results were not surprising, and thus provide a contrasting control condition for the phoneme-deviant results. The effects of task relevance and deviance position for the phoneme deviants are the most important part of the present results because the purpose of the experiment was to determine the contribution of phoneme processing to MMN generation. In spite of the complexity of these phoneme-deviant results, their implications with respect to the existence of the phoneme-specific MMN seem quite clear. Each major finding and its implications for MMN generation and modulation will be discussed in a separate section below.

A. Attention effects

In the present experiment, the listener had to attend to the stimuli in one ear in order to detect intensity /B/ or phoneme /g/ deviants in that ear. The MMN amplitude for unattended-ear deviants was only 51.3\% of that for attended-ear deviants, and this attention effect was independent of the deviant feature (intensity or phoneme), task relevance (target-type or nontarget-type) and deviance position (1st, 2nd or Both). In addition, there were no significant interactions that included attention, although the other factors did produce main effects and interactions among each other. Such a simple, non-interactive effect of attention suggests a simple mechanism.

Within the context of the model of MMN generation proposed by Näätänen (1992, 1995), there are three points at which attentional modulation could produce simple effects: (1) the output of the MMN generator, (2) the sensory input signals to the MMN generator, or (3) the memory trace of the standard stimulus. The results of many MMN-attention experiments would not be expected to discriminate among these alternatives. Modulation of the MMN output (alternative 1) would require a coupling between a process of identifying the stimulus as attended or unattended and the MMN modulator. This alternative is included in the model proposed here. Modulation of sensory input (alternative 2) might require early differentiation based on fundamental stimulus features like location or frequency band rather than a more complex stimulus
attribute like the direction of frequency glide. Such simple features could be effective because of the possibility of presetting a filter to enhance attended relative to unattended sensory input. Other, more complex features might also facilitate sensory gating, if they can serve as a basis for the perceptual organization of an ongoing sound stream (Bregman, 1990). For example, with subjects listening for information or shadowing one of two voices, Woods, Hansen and Hillyard (1984) demonstrated short-latency selective attention effects in ERPs for unrelated speech probes in the attended voice. Thus, alternative 2 has predictive value in that attention directed to highly complex stimulus attributes should not affect the MMN, whereas attention to more fundamental auditory features should produce attention effects.

Modulation of the memory trace of the standard stimulus (alternative 3) could provide a basis for the attended deviants to generate higher amplitude MMNs, if a more accurate or precise memory trace were made of the attended standard stimulus. Even if the comparison of each stimulus with the memory trace of its standard were completely preattentive, a better representation of the attended standard should generate larger difference signals for the attended deviants. In the present experiment, there was no difference between the attention effects on the phoneme (/g/) and intensity (/B/) deviants. This equality would be the expected result if the attention effect were based on a more precise representation of the standard /ibi/, because exactly the same standard would serve for intensity and phoneme deviants. In contrast, consider Alain and Woods (1997), where pure tone frequency-pattern and intensity were the different features in the attended and unattended ears. The pattern-feature MMN showed greater attentional modulation when intensity was the target than did the intensity MMN when frequency pattern was the target. A sharpened memory trace of frequency pattern would not require more precise intensity information, but a sharpened memory trace of standard-intensity tones would include precise frequency specification, consistent with the asymmetry in the attention effect that was found. Cowan (1995) provides an extensive framework for considering such effects of attention on the memory trace.

The memory-trace modulation model also can account for other differences in attention effects. Woods, Alho and Algazi (1992) (and Alho, Woods, Algazi and Näätänen, 1992) reported that the MMN was more sensitive to attentional manipulation when the frequency of the deviant (1064 or 1050 Hz) was closer to the frequency of the standard tone (1000 Hz) than when it was more distant (1500 Hz). Clearly, this would be the expected result if the effect of attention were to sharpen the representation of the standard stimulus: A more accurate memory trace of the 1000-Hz standard would be needed to differentiate it from the deviant when the frequency difference was less. Memory-trace modulation also might account for the lack of an attention effect in Alho, Woods and Algazi (1994). The auditory task in their attend-auditory conditions was to respond to only the more distant deviant (1300 Hz) and not to respond to the deviant (1050 Hz) that was closer to the standard (1000 Hz). If this task caused the subject to form a less precise memory trace of the standard stimulus than if the close deviant also had been a target (as in Woods, Alho and Algazi, 1992), then the memory-trace modulation model would be consistent with the absence of an attention effect.

In some studies of the effect of attention on the MMN, the attended and unattended stimuli were very different. Trejo et al. (1995), for example, measured pure-tone MMNs while subjects listened to a mixture of narrative speech and pure tones. MMNs to pure tone deviants were smaller when subjects attended to speech than when they attended to the tones. Alho et al. (1992) and Woods et al. (1992) also demonstrated that attention to a visual task produced smaller pure-tone MMNs relative to attention to the tones. Furthermore, although attention to vision produced smaller MMNs, the difficulty of the visual task did not affect the MMN amplitude. In the memory-trace
modulation model, attention to any different category of stimulus should eliminate attentional sharpening of the memory trace of the standard stimuli. Thus, the model is consistent with the reduction of the pure-tone MMN by attention to speech or a visual task. The insensitivity to visual-task difficulty also seems consistent with the model, as long as the effect of attention is characterized as a sharpening of the memory trace due to attention to those stimuli. In sum, the memory-trace modulation model seems to account for the results of the available MMN-attention studies more adequately than either of the other alternatives.

One final topic should be considered briefly before the discussion of attention effects is concluded. It is difficult to eliminate a possible N2b contribution to the present attention effect, or to the task-relevance effects to be discussed below (Näätänen, 1992; Alho et al., 1994; Alho, 1995). Two response properties can be used to distinguish the N2b from the MMN: (1) the N2b has a longer response latency than the MMN and (2) the N2b does not reverse polarity at the mastoid. Although neither of these properties provides a completely definitive means of separating N2b and MMN components, they can be used together to evaluate the N2b contribution to any particular effect. Briefly, the magnitude the effect can be measured in an earlier time interval, when the mastoid difference waves (relative to Nz) are positive. If a similar magnitude effect occurs in this earlier interval, then the effect can not be attributed to N2b contamination. In the case of the attention effect in the present experiment, the MMN reduction for deviants in the unattended channel was slightly larger in the earlier intervals: 58% versus 54% (in the original interval) for intensity deviants, and 48% versus 43% (in the original interval) for phoneme deviants. Thus, the early-interval analysis indicated that N2b contamination was not a major factor in this attention effect.

B. Intensity Deviants

The present results demonstrate that the deviance position and the task relevance of the deviant stimulus influenced the MMN for intensity deviants. (In the intensity deviants, the early-interval analysis indicated that neither the deviance-position nor the task-relevance effects could be attributed to N2b contamination.) In the case of deviance position, the intensity-deviant MMN showed the expected result: a larger response for the Both position and a smaller response for at least one of the single-position deviants. The single-position deviant that should show the smallest response, of course, is the one that proved more difficult to discriminate from the standard stimulus, i.e., the 2nd position intensity deviant. That is not to claim that the virtual equality of the MMN amplitudes for the 1st and Both positions was predicted, but merely that the 1st position MMN amplitude should have been in the amplitude range determined by the 2nd and Both position MMNs. Only an order of response amplitudes that clearly violated this physical/behavioral order could be taken as evidence that some other factor were controlling the MMN amplitude.

In the case of task relevance, the intensity deviant MMNs had greater amplitudes when the intensity feature was the target-type. It is important not to confuse the task relevance of the deviant feature with whether or not the deviant was the target. The target-type deviant was a target or a nontarget, depending on whether it was in the attended or unattended channel. Indeed, the attended nontarget-type deviant was more easily confused with the target. Both subjective report and false-alarm analysis showed that virtually all of the false-alarm responses were generated by attended nontarget-type deviants, not unattended target-type deviants, indicating that the task-relevance effect was not due to false-alarm or aborted-false-alarm behavioral responses.

Although there was no particular reason to expect that intensity-deviant MMN amplitudes would be larger for target-type compared to nontarget-type deviants, the result is not surprising and is consistent with the attention effect discussed above—target-type deviants and deviants in the
attended channel generate higher-amplitude MMNs. It is also interesting that the task relevance effect occurred independent of attention and independent of whether the deviance was in the 1st, 2nd or Both position. From this perspective, the effect of task relevance for the intensity deviants was similar to the effect of attention. It is as if the MMN generator can be set to focus on processing intensity because intensity defines the target. Indeed, it might make sense to adopt this processing-focus explanation, merely as attention focused along the intensity-feature dimension, if the effect of task relevance had not been so different in the case of the phoneme feature.

Few previous MMN studies have manipulated the task relevance of the deviant stimuli, even in the more limited context of targets versus nontargets in a single auditory channel. In a pioneering MMN study, Näätänen, Gaillard and Mäntysalo (1978) found no difference between MMNs for higher- and lower-frequency deviants (relative to an intermediate standard frequency), depending on which deviant frequency was the behavioral target. Although there are many differences between the methods of Näätänen et al. (1978) and the present study, the most striking is that their target and nontarget deviants were within the same feature (pure tone frequency), while the present experiment employed very different features—phoneme and intensity. Another salient difference is the behavioral task: counting the targets in Näätänen et al. (1978) and a speeded button press response to each target here. Two other studies have reported at least a trend in the direction of larger MMNs for target-type compared to nontarget-type deviants (as was found here for the intensity deviants). Nordby, Roth and Pfefferbaum (1988a) used pure tones in which the deviant stimulus either was a different frequency or was delivered at a shorter ISI. Target and nontarget frequency deviants produced the same MMN amplitudes, while the amplitude of the MMN for the ISI nontarget deviant at Fz was 2 μV less than that for ISI target. (This difference was not subject to any direct statistical tests and none of the tests performed revealed it as part of a significant interaction.) Alain and Woods (1997) studied the effect of attention on the MMN to tonal-pattern and intensity deviants, and found significant differences between the MMNs for targets and attended nontargets. Since no differential analysis of responses to unattended target-type and nontarget-type deviants was done to show whether or not a similar effect had occurred in the unattended channel, it is not known whether their effect was target specific or was a task-relevance effect as in the present study. Further study of the effect of task relevance on the MMN would be of interest to determine how top-down cognitive control can influence the MMN for both speech and nonspeech stimulus features.

C. Phoneme deviants

In contrast to the MMN amplitudes for intensity deviants, phoneme-deviant MMNs showed no significant amplitude variation with either deviance position or task relevance alone. For the phoneme feature, it was the interaction of task relevance and deviance position that was significant and also very interesting. (In the phoneme deviants, the early-interval analysis indicated that the deviance-position-by-task-relevance interaction could not be attributed to N2b contamination.) When the phoneme deviants were target-type, the MMN amplitudes for the different deviance positions were consistent with physical differences and detectability. When the phoneme deviants were nontarget-type, however, the MMN amplitudes for the different deviance positions were consistent with phonetic differences and clearly inconsistent with physical differences and detectability. This interaction indicates that the properties of the phonetic processing were adjusted to the demands of the behavioral task. When /g/ was the target-type, the focus of phonetic processing was on the presence or absence of the phoneme /g/, and the detectability of the /g/ largely determined the MMN amplitude. Only in the case of the 2nd-position phoneme deviant (/ibgi/), would the occurrence of the /g/ signal both the presence of the /g/ target and the presence of the
VCCV instead of the VCV. This would account for the relatively large 2nd-position MMN amplitude for the target-type /ibgi/, in spite of its lower detectability.

The situation was quite different when /B/ was the target. The phonetic processing had no particular focus in this case because there was no deviant-phoneme target—the target was just the standard phoneme at a higher intensity. Without a particular focus, complete automatic phoneme processing occurred and the full phonetic difference determined the MMN amplitude. Thus, the occurrence of the deviant /g/ and the possible occurrence of the deviant VCCV combined to contribute to the MMN, yielding the smallest MMN amplitude for the /igi/ where no VCCV component was present. The largest MMN amplitude was obtained for /igbi/, where the /g/ in the 1st position contributed to the MMN and then the /b/ in the 2nd position established the VCCV, which further contributed to the MMN. An intermediate MMN amplitude might be expected for the nontarget-type /ibgi/ because there was no deviance through the 1st position, and then the 2nd position established the presence of the /g/ and the VCCV simultaneously. There have been few other studies of double deviants (Czigler and Winkler, 1996; Levänen, Hari, McEvoy and Sams, 1993; Schröger, 1995; Winkler, Czigler, Jaramillo, Paavilainen and Näätänen, 1998) and none involving speech sounds or stimuli where one feature change produced double deviance, as in the present study. The data in all of the other studies showed at least a trend for double deviants to yield larger MMNs under some conditions. Only Schröger (1995, measuring MMNs) and Levänen et al. (1993, magnetic mismatch) used double deviants with temporally simultaneous deviation of two different features and both found double the mismatch response for the double deviant. In the other two studies, the onset of deviance for the two features was separated by at least 100 ms (longer than the time between the 1st and 2nd deviance positions in the present study) and the MMNs reflected less deviance additivity.

Independent of exactly how different components of deviance might combine to produce the MMN, the explanation of the present results in terms of /b/-/g/ and VCV-VCCV differences would yield one prediction. The task relevance effect on phoneme deviants should be minimal for the 2nd deviance position, as was seen in Fig. 6. Whether the processing was focused on detecting /g/ (as hypothesized in target-type mode) or was running on automatic (nontarget-type mode), the 2nd-position /g/ signaled its own presence as well as the presence of the VCCV. Thus, MMNs for target-type and nontarget-type 2nd-position phoneme deviants should be generated by the simultaneous occurrence of /g/ and VCCV deviation. The situation was different for the 1st-position /g/. In target-type mode, detecting the 1st-position /g/ inhibited or masked further processing and only the /g/ deviation contributed to the MMN; while in nontarget-type mode, the sequential occurrence of /g/ and VCCV deviations both contributed to the response.

D. Phoneme-specific MMN

A fundamental question in the present research concerns the nature of the MMN response to speech stimuli. Is the MMN response to deviant speech sounds primarily due to differences detected before or after speech-specific phoneme processing? If the phoneme MMN were due to differences in a general spectral-energy representation of auditory stimuli, prior to phoneme processing, then it could be used to measure the information content of that spectral-energy representation. Such a measure would be very useful in a theoretical context for understanding the nature and information content of the spectral-energy representation. It also would be of clinical value in the case of hearing impairment due to cochlear (or other subcortical) damage. On the other hand, if the phoneme MMN were due to differences in phoneme processing, then it should be of primary use in studying speech perception and speech-perception deficits.
Previous studies have suggested that the MMN may or may not be speech specific. Sharma, Kraus, McGee, Carrell and Nicol (1993) reported similar MMNs for speech sounds with similar physical differences, whether or not they were exemplars of different phonemes. Sharma and Dorman (1998) also recorded MMNs for different exemplars of the vowel /i/. Maiste, Wiens, Hunt, Scherg and Picton (1995) found no discontinuity in MMN amplitudes to indicate the location of the category boundary on the /ba/-/da/ continuum. In other studies using speech stimuli chosen to emphasize phonetic contributions to the MMN (Aaltonen et al., 1997; Dehaene-Lambertz, 1997; Näätänen et al., 1997; Tremblay et al., 1997), results suggest a phoneme-specific MMN. As noted in the Introduction, however, even these MMNs could have resulted from either pre- or post-phonemic differences. The same conclusion would apply to the demonstration of the McGurk effect in the magnetic analog to the MMN (Sams, Aulanko, Hämäläinen, Hari, Lounasmaa, Lu and Simola, 1991), as long as the McGurk effect could be due to visual modification of pre-phonetic auditory information.

In contrast, it seems more difficult to explain the present results for the nontarget-type phoneme deviants on the basis of strictly pre-phonemic differences. In terms of the spectral-energy representation of the phoneme deviants, the Both position differed maximally from the standard—as much as the 1st- and 2nd-position differences combined. Furthermore, the behavioral results and the MMN amplitudes for target-type phoneme deviants were generally consistent with the idea that differences in spectral-energy representations determine both behavioral discrimination and MMN amplitude. The MMN response to the 2nd position target-type phoneme deviant was larger than expected, but that single point would not be very important without considerable other evidence for the phoneme-specific MMN. This other evidence is provided by the MMN amplitudes for nontarget-type phoneme deviants. Here, the MMN amplitudes were clearly inconsistent with the physical magnitude of stimulus differences and with the behavioral performance. Furthermore, they showed the pattern of the perceived-phoneme differences, where the combination of the presence of the /g/ and the phonetic doubling for the central consonant (VCV to VCCV) can convert a smaller physical/behavioral difference into a larger phonetic difference. The hemispheric asymmetry in the deviance position effect further supported a phonetic source for this MMN, as does the larger than expected MMN for the 2nd position phoneme deviant even when it is the target-type. It is not apparent how an explanation that relies on acoustic differences and excludes the phonetic VCV-VCCV distinction could account for these results.

Connolly and Phillips (1994; Connolly, Phillips and Forbes, 1995) have reported a phonetic mismatch negativity (PMN) in a paradigm that included meaningful sentences rather than repeated standard syllables with rare phoneme-deviant syllables. The PMN occurred in response to a phonetic mismatch between the initial phoneme of an expected word and that of a substituted, semantically appropriate word with a different initial phoneme. From one perspective, the only difference between the PMN and the MMN of the present study is in how the "expected" phoneme is defined—by linguistic context or by repetition. Winkler, Cowan, Csépe, Czigler and Näätänen (1996) demonstrated that only a single exemplar of a pure-tone standard was needed to reinstate its memory trace across a 9.5 s gap filled with random tones. Perhaps, in the case of meaningful speech, the mere expectation of a phoneme may be sufficient to load the memory trace. Under normal circumstances, such a process would facilitate the understanding of speech by reducing the time spent on phoneme identification.

V. CONCLUSIONS

Attention can have a major impact on MMN amplitudes, but any explanation must be able
to produce a simple, non-interactive increase in the MMN for attended stimuli. A model in which attention improves the memory trace of the standard stimulus would account for these results and also offer an explanation for the many differences in attention effects seen in previous studies.

The task relevance of a deviant feature also affects the MMN amplitude, but this effect and the presence or absence of interactions was different for intensity and phoneme deviant features. These findings indicate that MMN generation is much more complex than a simple pre-attentive detection of physical deviance.

The magnitudes of perceived phoneme differences may control MMN amplitudes, even when these magnitudes are inconsistent with physical differences and discrimination differences measured in the same listeners during the same experiment. The existence of the phonetic MMN undermines the use of the MMN in measuring the spectral-energy representation of speech sounds in the central auditory pathways and supports its use in studies of speech perception.

Acknowledgments

The work described above was supported by the Department of Veterans Affairs and by NS32893. The authors thank Lydia Peters for her technical assistance in data collection, and Claude Alain, Kimmo Alho, Keith Ogawa and Steven Thomas for their helpful comments.

Textual Footnotes

1. Of the initial subject pool of 28, five failed to reach criterion performance in the training session and another six failed to complete both recording sessions. In the process of ERP averaging, data of four of the remaining 17 subjects showed excessive artifact rejections, leaving the final 13 subjects whose data were included in the analysis. Although less than 50% of the initial subjects produced usable data, there should be no concern about this producing a biased sample population because the loss at the two critical points (5 of 28 and 4 of 17) is not excessive for those situations. All of the subjects were recruited from outside of the laboratory, none knew the purpose of the experiment, and none had previous experience in auditory, or more than one visual, ERP experiment.

2. There is some ambiguity concerning the appropriate terminology for these stimuli because phonetic and physical representations seem to conflict. In phonetic terms, /ibi/ and /igi/ are VCVs and should not be called /ibbi/ and /iggi/, while /ibgi/ and /igbi/ are VCCVs and should be so called. In physical terms, /ibi/ and /igi/ differ during both VC and CV formant transitions, suggesting that /ibbi/ and /iggi/ might be used to emphasize this physical difference and the fact that both /ibgi/ and /igbi/ are physically intermediate between /ibi/ and /igi/. The phonetic terminology will be used here because the purpose of the experiment was to determine the role of phonetic differences in MMN generation and because the phonetic terminology makes it easier to explain the phonetic hypothesis that /ibgi/ and /igbi/ deviants might generate larger MMN responses than the /igi/ deviant. For consistency, similar terminology will be used for the intensity deviants (/ibBi/, /iBbi/ and /iBi/).

3. Indeed, the fact that the onset of deviance occurred at 60 ms (Both and 1st positions) or 130 ms (2nd position), indicates that a few of the durations of significant negativity must include initial points of spurious significance (type I errors) that happen to be adjacent to the MMN. Note, for example, that the duration of significant negativity for the attended target-type /iBbi/ begins at 80 ms—only 20 ms after deviance onset. Comparing the attended target-type waveforms for /iBbi/ and /iBi/,
which are plotted in Fig.9A for another purpose, supports the suggestion that the initial negativity in the difference wave for /iBbi/ is not a part of the MMN response for that stimulus. The difference waves for /iBbi/ and /ibi/ are very similar in the range 150-300 ms, while the difference wave for /iBbi/ has a seemingly independent negativity in the range 80-150 ms.

4. Näätänen et al. (1993) found a similar attentional modulation of the pure-tone frequency MMN when the target was defined by a difficult-to-detect intensity deviant. Since virtually no intensity MMN was found unless the intensity deviant was the target, there was no measurement of the attentional modulation of the intensity MMN when the frequency deviant was the target. The lack of MMN for difficult-to-detect nontarget intensity deviants is consistent with Woldorff et al. (1991). Näätänen et al. (1993) attribute the MMN-like component in the intensity-target difference waves to the N2b component generated by target stimuli.

5. The time intervals used for the early-interval analysis were 150-250 ms for Both- and 1st-position intensity and phoneme deviants, 200-300 ms for 2nd-position intensity deviants, and 220-320 ms for 2nd-position phoneme deviants. All but four of the 24 mastoid difference waves (corresponding to the conditions for the 24 Fz difference waves of Figs. 7 and 9) showed a positive average voltages in these early intervals and these four exceptions all showed a positive voltage shift within the interval, but not sufficiently so to result in a positive average voltage over the entire interval.

References


Table 1: The deviant stimuli used in the present experiment. The standard stimuli were the VCV /ibi/. Listed here are the deviance position, the time of deviance onset relative to the beginning of the initial vowel (time = 0 ms) and the notation for each phoneme and intensity deviant.

<table>
<thead>
<tr>
<th>Deviance Position</th>
<th>Onset of Deviance (ms)</th>
<th>Phoneme Deviant</th>
<th>Intensity Deviant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both</td>
<td>60</td>
<td>/igi/</td>
<td>/iBi/</td>
</tr>
<tr>
<td>1st</td>
<td>60</td>
<td>/igbi/</td>
<td>/iBbi/</td>
</tr>
<tr>
<td>2nd</td>
<td>130</td>
<td>/ibgi/</td>
<td>/ibBi/</td>
</tr>
</tbody>
</table>
Table 2: Mean RTs, target-detection hit rates and false alarm responses (FAs) for deviant-phoneme and deviant-intensity targets in each deviance position. RTs were measured from the onsets of deviance given in Table 1. Note that either the phoneme or intensity deviants in one ear were the behavioral targets, independent of the position of the deviance within the consonant portion of the VCV.

<table>
<thead>
<tr>
<th>Position</th>
<th>RT (ms)</th>
<th>Hit Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phoneme</td>
<td>Intensity</td>
</tr>
<tr>
<td>Both</td>
<td>389</td>
<td>375</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>402</td>
<td>401</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>415</td>
<td>401</td>
</tr>
<tr>
<td>All</td>
<td>402</td>
<td>392</td>
</tr>
<tr>
<td>FAs (number per block of 800 stimuli)</td>
<td>2.38</td>
<td>1.35</td>
</tr>
</tbody>
</table>
Table 3: The durations of significant difference-wave negativity for phoneme and intensity deviant features under all combinations of the other three factors (attention, deviance position and task relevance). Mean voltages of the difference waves were measured in 20-ms intervals for each subject and the significance of the difference from zero was determined (ANOVA). Only durations with at least three consecutive significant 20-ms intervals, including at least two with $p < 0.01$, are listed here. The time intervals are defined relative to stimulus onset, rather than to the onset of the deviance, to facilitate the identification of significant portions of the differences responses in Figs. 7 and 9. (To convert to intervals relative to deviance onset, subtract 60 ms from Both- and 1st-position values and 130 ms from 2nd-position values.)

<table>
<thead>
<tr>
<th>Deviant Stimulus</th>
<th>Attended Target-Type</th>
<th>Attended Nontarget-Type</th>
<th>Ignored Target-Type</th>
<th>Ignored Nontarget-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iBi/</td>
<td>160-280</td>
<td>140-280</td>
<td>160-300, 560-720</td>
<td>200-260</td>
</tr>
<tr>
<td>/iBbi/</td>
<td>80-280</td>
<td>140-280</td>
<td>140-280</td>
<td>180-260</td>
</tr>
<tr>
<td>/ibBi/</td>
<td>160-360</td>
<td>220-360</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>/igi/</td>
<td>120-340</td>
<td>200-300</td>
<td>200-360</td>
<td>200-300</td>
</tr>
<tr>
<td>/ibgi/</td>
<td>240-440</td>
<td>220-400</td>
<td>280-380</td>
<td>ns</td>
</tr>
</tbody>
</table>
Figure 1: Waveform (A) and spectrogram (B) of the standard VCV stimulus, /ibi/. The onset of the initial vowel and the offset of the final vowel were shaped with a 10 ms linear ramp. Steady-state vowels were present in the intervals 10-60 ms and 160-210 ms. The VC formant transitions occurred at 60-90 ms, the stop-consonant gap occupies 90-130 ms and the CV formant transitions occurred at 130-160 ms. Deviant stimuli differed from standards only in the VC and/or CV formant-transition intervals.
Figure 2: A 6-second example of the experimental paradigm when the target was phoneme-feature deviance in the right ear. Asterisks indicate target sounds, a Both-position and a 1st-position phoneme deviant—the 2nd-position phoneme deviant /ibgi/ in the right ear was also one of the targets, but it did occur in this 6-second interval. Note the example begins with a standard left-ear sound, switches to a right-ear nontarget 1st-position intensity deviant, then three left-ear sounds, the third of which is a target-type Both-position phoneme deviant, etc. In other blocks, the right-ear intensity deviants or the left-ear phoneme or intensity deviants were the target sounds.
Figure 3: ERPs for standard and deviant stimuli at Fz averaged across all subjects. Each panel includes ERPs for a standard stimulus and for the intensity and phoneme deviants that were presented in the same ear under the same condition. Here, the data are collapsed across ear of presentation and deviance position. Responses are shown for stimuli in the attended ear in A and C and for stimuli in the unattended ear in B and D. The ERPs fail to show the clear pattern of components that would be expected for brief-tone-burst or click stimuli, probably because of their long duration and the amplitude and formant-frequency variation within the stimuli. Expected negative and positive difference components, however, are apparent when the waveforms for deviant and standard stimuli are compared. The ERPs are plotted relative to the average level in the baseline interval (the 200 ms prior to stimulus onset). Time 0 is stimulus onset. The voltage scale was chosen to illustrate the negative differences between deviant and standard ERPs, putting the large positive (P3) response to target stimuli off the bottom of the negative-up voltage ordinate.
Figure 4: The MMN in attended and ignored conditions for nontarget-type phoneme deviants averaged across all subjects. The difference waves are plotted relative to the average baseline level (from -200 to 0 ms) and the data are pooled over ear of stimulation and deviance position. Note the fronto-central scalp distribution and the larger MMN amplitude when stimuli were attended. The responses at Fz and Cz are expanded in the call-out box to permit more detailed observation; in these responses combined across deviance position, the peak of the MMN occurs at about 290 ms.
Figure 5: The MMN in attended and ignored conditions for nontarget-type intensity deviants. Otherwise the same as Fig. 4. Note the fronto-central scalp distribution and the larger MMN amplitude when stimuli were attended, similar to the phoneme data of Fig. 4. The peak of the MMN occurs at about 240 ms, or about 50 ms earlier than for phoneme deviants.
Figure 6: The interaction of deviance position and task relevance for phoneme deviants. The mean amplitude of the MMN at Fz is plotted for each deviance position for target-type deviants (cross-hatched bars) and nontarget-type deviants (solid bars). See text for further discussion.
Figure 7: Average phoneme MMNs at Fz for each deviance position (Both in solid, 1st in dotted and 2nd in dashed lines) for: (A) attended target-type deviants, i.e. targets; (B) attended nontarget-types; (C) unattended target-types; and (D) unattended nontarget-types. Note the same interaction between deviance position and task relevance independent of attention condition. See text for further discussion.
Figure 8: The effect of deviance position and task relevance for intensity deviants. The mean amplitude of the MMN at Fz is plotted for each deviance position for target-type deviants (cross-hatched bars) and nontarget-type deviants (solid bars). Note the lack of interaction between deviance position and task relevance: Whether the intensity deviant is target- or nontarget-type, the 1st and Both deviance positions produced equally large MMNs while the 2nd deviance position produced a much smaller response. Compare with the interaction seen for phoneme deviants in Fig. 6. (In comparing Figs. 6 and 8, note the different MMN-amplitude scales; the 2nd-position intensity-deviant responses were smaller, while the 1st- and Both-position target-type intensity-deviant responses were larger, than any of the phoneme-deviant responses.)
Figure 9: Average intensity MMNs at Fz for each deviance position (Both in solid, 1st in dotted and 2nd in dashed lines) for: (A) attended target-type deviants, i.e. targets; (B) attended nontarget-types; (C) unattended target-types; and (D) unattended nontarget-types. In contrast to the phoneme MMN of Fig. 7, intensity MMNs show essentially the same deviance-position effect in all four panels. The task-relevance main effect is also apparent, with target-type MMNs greater than nontarget-type MMNs.
Figure 10: The interaction of deviance position and electrode-site hemisphere for phoneme deviants. The mean amplitude of the MMN is plotted for each deviance position for right-hemisphere (RH, cross-hatched bars) and left-hemisphere (LH, solid bars) electrode sites. Data were collapsed across attention condition and task relevance, with F4, FC6, C4, CP6 and P4 averaged to obtain the RH response amplitudes and F3, FC5, C3, CP5 and P3 averaged to obtain the LH response amplitudes.