J Am Acad Audiol 15:47–66 (2004)

Processing of English Words with Fine Acoustic Contrasts and Simple Tones: A Mismatch Negativity Study

Catharine M. Pettigrew*
Bruce M. Murdoch*
Joseph Kei*
Helen J. Chenery*
Ravi Sockalingam* ∫
Curtis W. Ponton†
Simon Finnigan‡
Paavo Alku§

Abstract
The purpose of this study was to compare the robustness of the event-related potential (ERP) response, called the mismatch negativity (MMN), when elicited by simple tone stimuli (differing in frequency, duration, or intensity) and speech stimuli (CV nonword contrast /de:/ vs. /ge:/ and CV word contrast /deI/ vs. /geI/). The study was conducted using 30 young adult subjects (Groups A and B; n = 15 each). The speech stimuli were presented to Group A at a stimulus onset asynchrony (SOA) of 610 msec and to Group B at an SOA of 900 msec. The tone stimuli were presented to both groups at an SOA of 610 msec. MMN responses were elicited by the simple tone stimuli (66.7%–96.7% of subjects with MMN "present," or significantly different from zero, $p < 0.05$) but not the speech stimuli (10% subjects with MMN present for nonwords, 10% for words). The length of the SOA (610 msec or 900 msec) had no effect on the ability to obtain consistent MMN responses to the speech stimuli. The results indicated a lack of robust MMN elicited by speech stimuli with fine acoustic contrasts under carefully controlled methodological conditions. The implications of these results are discussed in relation to conflicting reports in the literature of speech-elicited MMNs, and the importance of appropriate methodological design in MMN studies investigating speech processing in normal and pathological populations.

Key Words: Consonant-vowel tokens, event-related potential, mismatch negativity, simple tones, stimulus onset asynchrony, words

Abbreviations: ERP = event-related potential; FA = false alarms in behavioral task; HR = hit rate for behavioral task; MMN = mismatch negativity; MMNm = magnetic mismatch negativity; SNR = signal-to-noise ratio; SSG = semisynthesized speech generation

*Department of Speech Pathology and Audiology, University of Queensland, QLD, Australia; †Neuroscan Labs, El Paso, Texas; ‡School of Psychology and the Centre for Magnetic Resonance Imaging, University of Queensland, QLD, Australia; §Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland; ∫ School of Human Communication Disorders, Dalhousie University, Halifax, Canada

Reprint requests: Catharine M. Pettigrew, Department of Speech Pathology and Audiology, University of Queensland, QLD, 4072, Australia; Phone: +61-7-3365-6149; Fax: +61-7-3365-1877; E-mail: c.pettigrew@shrs.uq.edu.au

Sumario

El propósito de este estudio fue comparar la fortaleza de un potencial relacionado con el evento (ERP) conocido como "negatividad desigual" (mismatch negativity: MMN), cuando se genera por medio de estímulos tonales simples (con diferencias en frecuencia, duración o intensidad), o con estímulos de lenguaje (partículas en contraste CV /de:/ vs. /ge:/ y palabras en contraste CV /de/ vs. /ge/. El estudio fue conducido utilizando 30 sujetos adultos jóvenes (Grupos A y B; n = 15 cada grupo). Los estímulos del lenguaje fueron presentados al Grupo A con una asincronía inicial del estímulo (SOA) de 610 mseg, y al Grupo B con una SOA de 900 mseg. Los estímulos tonales se presentaron a ambos grupos con un SOA de 610 mseg. Las respuestas MMN fueron generadas por estímulos tonales simples (66.7%-96.7% de los sujetos con MMN "presente" o significativamente diferente de cero, $p < 0.05$) pero no ante los estímulos del lenguaje (10% de los sujetos con MMN presentes para partículas no lingüísticas, 10% para palabras). La longitud del SOA (610 mseg o 900 mseg) no tuvo ningún efecto sobre la capacidad de obtener respuestas MMN consistentes con el estímulo de lenguaje. Los resultados indicaron una falta de respuesta MMN robustas, producidas por estímulos de lenguaje con contrastes acústicos finos bajo condiciones metodológicas cuidadosamente controladas. Las implicaciones de estos resultados se discuten con relación a los reportes conflictivos hallados en la literatura sobre las MMN generadas con lenguaje, y sobre la importancia de un diseño metodológico apropiado en estudios de MMN que investigan el procesamiento del lenguaje en poblaciones normales y patológicas.

Palabras Clave: fichas consonante-vocal, potencial relacionado con el evento, negatividad desigual, tonos simples, asincronía inicial del estímulo, palabras

Abreviaturas: ERP = potencial relacionado con el evento; FA = falsas alarmas en tareas conductuales; HR = tasa de aciertos para tareas conductuales; MMN = negatividad desigual; MMNm = negatividad desigual magnética; SNR = tasa de relación señal/ruido; SSG = generación de lenguaje semi-sintético

The understanding of spoken language and the ability to communicate effectively require accurate neural processing of language-specific speech sounds (Kraus, McGee, Carrell, and Sharma 1995; Tallal et al, 1996). Like most cognitive processes, linguistic processing is associated with brain activity that can be measured by electroencephalography (EEG) (Kutas and Van Petten, 1994). In particular, event-related potentials (ERPs) are measures of brain activity with a high temporal resolution that can be time-locked to neural events during language processing (Hillyard, 2000; Kutas, 2000). The mismatch negativity (MMN) is a fronto-centrally distributed ERP component that is elicited by any discriminable auditory change (Näätänen et al, 1978). For example, an MMN is elicited when a tone changes in frequency, duration, or intensity, or a phoneme is replaced by another phoneme (Näätänen and Escera, 2000). The MMN is elicited in response to a repetitive sequence of auditory stimulation using an oddball paradigm, and occurs about 100–200 msec after the onset of the "deviant" stimulus. Most importantly, the MMN response can be elicited in the absence of attention and does not rely on behavioral response (Näätänen et al, 1978, 1993). The MMN has been described as an objective physiological measure of central sound representations, including those for speech, in the human brain (Näätänen and Alho, 1997; Näätänen and Winkler, 1999; Näätänen, 2001).

Previous studies have shown that the MMN can index auditory discrimination of phonemes (Aaltonen et al, 1993; Titova and Näätänen, 2001), consonant-vowel (CV) syllables (Sams et al, 1990; Kraus et al, 1992; Aulanko et al, 1993; Kraus, McGee, Micco et al, 1993; Kraus et al, 1996), and pseudowords (Ceponiene et al, 1999). These observations indicate that the MMN can reflect the initial stages of acoustic-phonetic processing in the central nervous system. Further studies have used cross-linguistic designs to show that the MMN is also capable of indexing the accuracy of phonetic representations in the brain, that is, the detection of phonemes specific to the native language, in both adult and child populations (Dehaene-Lambertz, 1997; Näätänen et al, 1997; Cheour et al, 1998; Winkler, Lehtokoski et al, 1999; Sharma and Dorman, 2000).

More recently, Pulvermüller et al (2001) recorded MMNm (magnetic mismatch negativity) responses in Finnish subjects to the same spoken syllable completing a Finnish word or a pseudoword. The authors found that the MMNm to each syllable was larger when it completed a word rather than a pseudoword. As this MMNm enhancement did not occur in foreign subjects who did not know any Finnish, the authors proposed that the results provide evidence for the presence of memory traces, or neural representations, for individual spoken words in the human brain.

However, unlike the consistent findings among studies investigating speech processing at smaller linguistic levels, the few studies that have further investigated word-related processing using the MMN have demonstrated conflicting results, a problem mainly attributable to distinct differences in methodological design and data analysis procedures. The inconsistency of the findings reported by these studies raises the question of whether or not the MMN can reliably be elicited by speech stimuli at the more acoustically and linguistically complex level of words, when methodological design and data analysis procedures are carefully controlled.

More specifically, three recent studies by Korpilahti et al (2001), Wunderlich and Cone-Wesson (2001), and Shtyrov and Pulvermüller (2002) have contrasted words and pseudowords in a variety of separate conditions (word-word; pseudoword-pseudoword, word-pseudoword). Korpilahti et al (2001) investigated cortical ERPs in children in response to complex tones, naturally spoken words (Finnish), and pseudowords, in separate conditions. In addition, each stimulus type contained equal acoustical elements. The authors found that all three stimulus types elicited MMN responses, with the tones eliciting a greater MMN response than the words. However, the authors reported no significant difference in the mean amplitudes of the word and pseudoword mismatch responses.

Wunderlich and Cone-Wesson (2001) studied the effects of stimulus frequency and complexity on components of the adult cortical auditory evoked potential, and included syllables (consonant-vowel tokens, or CVs) and English words (consonant-vowel-consonant tokens, or CVCs), differing only in the place of articulation of the initial consonant (/bae/ vs. /dae/ and /baed vs. /daed/ respectively). In contrast to Korpilahti et al (2001), the authors were unable to obtain consistent MMN responses to the easily discriminable synthesized speech stimuli, and noted that the words were no more effective in eliciting the MMN response than the syllables.

More recently, Shtyrov and Pulvermüller (2002) used monosyllabic English words and pseudowords to further investigate the nature of the word-related enhancement of the MMN response reported by Pulvermüller et al (2001). Shtyrov and Pulvermüller (2002) found that the MMNs elicited by word deviants in both pseudoword and word standard conditions were larger than those elicited by pseudoword deviants presented among words. Furthermore, the MMN responses to word deviants did not significantly differ when presented among pseudoword or word standards, indicating that the presence of an enhanced MMN response to real word deviants appears to be independent of the lexicality of the standard. It is important to note, however, that although the difference waveform (MMN) responses in each of the conditions used in the study by Shtyrov and Pulvermüller (2002) are apparent visually, the MMN amplitudes are all less than a microvolt.

It is likely that methodological differences such as the subject population chosen, the language used, and the nature in which the speech stimuli were generated may have contributed to the conflicting results of the studies by Korpilahti et al (2001), Wunderlich and Cone-Wesson (2001), and Shtyrov and Pulvermüller (2002). For example, both Korpilahti et al (2001) and Shtyrov and Pulvermüller (2002) employed word and pseudoword stimuli that were naturally spoken by a female native speaker, but only Shtyrov and Pulvermüller (2002) attempted to make the stimuli as acoustically similar as possible and ensure identical standard-deviant contrasts by recording multiple repetitions of the stimuli, cross-splicing phonemes, and normalizing the stimuli to have the same peak sound energy. It is well-known that the MMN is particularly sensitive to acoustic differences between the standard and deviant stimulus (Kraus, McGee, Carrell, and Sharma, 1995; Näätänen, 2001; Shtyrov and Pulvermüller, 2002). Therefore, it is possible that the larger amplitude responses demonstrated by Korpilahti et al (2001) to the speech stimuli (4–5 μV) may be at least partially due to acoustic differences between the stimuli. Furthermore, similar to most previous MMN studies, Korpilahti et al (2001) subtracted the standard stimulus response from the deviant stimulus response, thereby obtaining an MMN response that did not exclude physical (stimulus-related) differences that could be attributable to differences in neural encoding of the two stimuli (Kraus, McGee, Carrell, and Sharma, 1995; Dalebout and Stack, 1999). Such a subtraction method assumes that other ERP components (e.g., N1) are not affected by any physical differences between the deviant and standard stimuli, which is not always the case (Schröger, 1998), and makes it difficult to interpret the MMN response (such as a word-related enhancement) in terms of the purely contextual differences between the standard and deviant stimuli. Although Shtyrov and Pulvermüller (2002) also used the deviant-minus-standard method to compute the MMN responses in their study, the physical differences between the deviant and standard stimuli in all three conditions were more carefully controlled, enabling the authors to make reliable comparisons between the responses in the three conditions.

In contrast, Wunderlich and Cone-Wesson (2001) controlled for all physical differences between the standard and deviant stimuli in the calculation of the difference waveforms by using a counterbalanced oddball paradigm and subtracting the averaged response to the deviant stimulus when it acted as a standard, from the averaged response to the deviant when it acted as a deviant (reverse contrasts). In addition, the authors used synthesized speech stimuli, which allowed them to precisely control for acoustic differences between the standard and deviant stimuli, and ensure identical acoustic contrasts in the syllable and word conditions. Despite taking these careful measures, Wunderlich and Cone-Wesson (2001) were unable to obtain robust MMN responses to the speech stimuli in their study. Unlike Korpilahti et al (2001) and Shtyrov and Pulvermüller (2002), however, Wunderlich and Cone-Wesson (2001) used speech stimuli of different acoustic structure across conditions (CV syllables vs. CVC words). This may have been a confounding variable, and the authors suggested that further studies investigate MMN responses to English words and pseudowords that use the same acoustic structure (e.g., CV pseudowords and CV words). Although Shtyrov and Pulvermüller (2002) used English words and pseudowords of the same acoustic structure (CVC), their study did not include a pseudoword-pseudoword condition similar to Korpilahti et al (2001) and Wunderlich and Cone-Wesson (2001).

The aim of the present study was to reconcile the conflicting results of these recent studies investigating MMN responses to English word stimuli, by determining whether the MMN could be elicited by CV nonwords and CV words in separate conditions using carefully controlled methodological design and data analysis procedures. First, the MMN responses of the same group of subjects to simple tone stimuli, which are known to provide reliable and robust MMNs (Sams et al, 1985; Näätänen, Paavilainen, Alho et al, 1989; Näätänen, Paavilainen, and Reinikainen, 1989; Giard et al, 1990; Näätänen, 1990; Paavilainen et al, 1991; Näätänen, 1992; Lyytinen et al, 1992; Joutsiniemi et al, 1998), were investigated in order to validate the methods of data collection and analysis used in the present study. Second, semisynthesized CV speech tokens (Alku et al, 1999) were generated based on the suggestion of Wunderlich and Cone-Wesson (2001) to employ CV syllable speech tokens of similar acoustic structure but differing in meaningfulness, and were presented in separate nonword and word conditions. It was hypothesized that a larger MMN response would be elicited by the meaningful CV stimulus contrast d/g (within /deI/-/geI/, "day"-"gay") than the acoustically identical, but nonmeaningful, CV stimulus contrast d/g (within /de:/-/ge:/), thereby reflecting the activation of memory traces for words in the cortex. Last, the present study aimed to determine whether a shorter stimulus onset asynchrony (SOA) of 610 msec could be applied to oddball paradigms containing speech stimuli, rather than the more commonly used longer SOAs (e.g., 900 msec) (Shtyrov et al, 1998; Shtyrov, Kujala, Ilmoniemi et al, 2000). It was hypothesized that there would be no difference in the MMN responses using either SOA. The applicability of a shorter SOA would be extremely useful in the reduction of testing times and is therefore an important aspect of the methodological design. This investigation into the most appropriate and effective methodological design for eliciting word-related MMNs is essential to the establishment of the MMN technique as a valuable tool in the clinical diagnosis of automatic auditory processing deficits at the word level.

METHODS

Subjects

Thirty healthy young adults volunteered to participate in the study. The subjects were randomly divided into two groups, Group A (n = 15, four males) and Group B (n = 15, three males). Subjects in Group A were aged between 17–25 years (mean 20.8 ± 1.8) and were presented with the speech stimuli at a stimulus onset asynchrony (SOA) of 610 msec. Subjects in Group B were aged between 19–38 years (mean 22.5 ± 4.8) and were presented with the speech stimuli at an SOA of 900 msec. Both groups were presented with the tone stimuli at an SOA of 610 msec. All subjects who participated in the ERP recordings were native speakers of English and considered the CV syllables /deI/ and /geI/ to be real words, but not the CV syllables /de:/ and /ge:/. All subjects gave their informed consent prior to testing and had normal hearing, as defined by pure-tone thresholds of 15 dB HL or better from 500–4000 Hz.

Stimuli

All subjects were presented with simple tone stimuli and speech stimuli. The deviant tone stimuli differed from the standard tone stimuli in frequency, duration, or intensity. The speech stimuli included a nonword CV syllable contrast (/de:/ vs. /ge:/) and a real word CV syllable contrast (/deI/ vs. /geI/).

The features of the simple tones were adapted from Paavilainen et al (1991) and were generated using Neuroscan®, STIM software. The standard tones were 600Hz, 75 dB SPL sinusoidal tone bursts with 50 msec duration, including a 5 msec rise/fall time. The frequency deviant stimuli were 650 Hz tones; the duration deviant stimuli were 20 msec tones; and the intensity deviant stimuli were presented at 60 dB SPL. Two blocks of 1008 simple tone stimuli were presented to each subject using a multiple deviant oddball paradigm, which enabled the testing time to be reduced without affecting the amplitude of the MMN response to each deviant stimulus type (Deacon et al, 1998). Each of the deviant stimuli (frequency, duration, and intensity deviants) had a 12.5% probability of occurrence, resulting in a total of 252 sweeps obtained in response to each deviant stimulus. The stimuli were presented with a constant SOA of 610 msec and in a pseudorandom fashion, so that no fewer than two standards were presented between deviant stimuli. In addition, each deviant stimulus was presented alone in a single block of 300 stimuli to obtain responses for the "deviant alone" condition.

The speech stimuli consisted of four CV stimuli (the nonwords /de:/ "de" and /ge:/ "ge," and the words /deI/ "day" and /geI/ "gay") and were created using the semisynthesized speech generation (SSG) technique (Alku et al, 1999). SSG makes it possible to synthesize natural sounding, yet controllable, speech stimuli by exciting an artificial vocal tract model with a real excitation of the human voice production mechanism, the glottal waveform. The formant values required for synthesis of the speech stimuli were computed with SSG from natural utterances produced by a native male speaker of Australian English (see Table 1 for values of the lowest four formants).

All the stimuli were synthesized using the same glottal excitation waveform, which implies that the fundamental frequency was the same across all of the stimuli and the acoustic differences between the stimuli were caused solely by the formant settings. More specifically, the CV stimuli were created so that the formant settings shifted continuously in time from one phoneme to another. For example, the start of the CV word "day" was in the formant settings of the phoneme /d/. After 45 msec, the settings were of the phoneme /e/, after which there was a time span of 91 msec during which the settings shifted to the phoneme /I/. During the last part of the word (duration 168 msec), the formant settings were in the positions of the phoneme /I/. However, in order to make the vowel sound more natural, the formant

Table 1. Center Frequencies of the Lowest Four Formants of the Speech Stimuli

	Formant Frequencies (Hz)			
Phoneme	**F1**	**F2**	**F3**	**F4**
/e/	689	1615	2562	3854
$/I/_1$	517	2046	2562	4000
$/I/_2$	409	2239	2606	4350
/d/	474	1701	2778	3941
/g/	495	1938	2498	3704

Note: The phonemes $/I/_1$ and $/I/_2$ are representatives of the same phoneme /I/ in English.

settings of /I/ were shifted from one setting (phoneme $/I/_1$ in Table 1) to another (phoneme $/I/_2$ in Table 1). For the word "gay," the lengths of the segments were the same as those for the word "day." For the two nonword CV stimuli, the length of the first segment (i.e., the shift from /d/ to /e/ for the nonword "de" and the shift from /g/ to /e/ for the nonword "ge") was also 45 msec. However, during the remainder of the nonwords, the settings were in a constant position (i.e., the phoneme /e/). Since all four stimuli began with a voiced plosive, a low amplitude segment of 28 msec was added to the beginning of each word. This segment, called the voice bar, reflects the fundamental frequency of phonation with no formant structure (Kent and Read, 1992). Thus, the total duration was 332 msec for all the stimuli.

The fundamental frequency (F_0) of the stimuli changed over time according to the intonation used by the speaker, as the glottal excitation used in the synthesis of the stimuli was computed from a natural utterance. The value of F_0 was 130 Hz at the beginning of the stimuli and was 105 Hz at the end of the stimuli. Finally, normalizing the energies of the stimuli equalized sound intensity. All speech stimuli were presented at an intensity level of 75 dB SPL.

The CV syllables /de:/ and /ge:/ were reported to be "nonwords," and the CV syllables /deI/ and /geI/ were reported to be "real words" by ten native speakers of Australian English, prior to the commencement of testing (these native speakers did not participate in the ERP recordings). As the procedures used to generate the CV syllables /de:/ and /ge:/ involved the truncation of the diphthong vowel within the CV words /deI/ and /geI/, and the subsequent lengthening of that vowel to match the stimulus duration of the CV words, the resultant vowel within the CV syllables comprised formant values that do not easily fall within the range of a traditional singular vowel of English (Kent and Read, 1992). More specifically, the first and second formant values for the pseudovowel /e:/ bordered the outer formant value possibilities for the English vowels "u" in "up," "e" in "herd," and "a" in "at." As a result, the CV syllables /de:/ and /ge:/ were not considered to follow the phonotactic rules of English and were therefore labelled as "nonwords" rather than "pseudowords."

In the CV speech stimulus conditions, subjects were presented with three blocks of 800 stimuli per test condition, using a single contrast oddball paradigm. In the CV nonword condition, the stimulus /de:/ acted as the standard (87.5% probability), and the stimulus /ge:/ acted as the deviant (12.5% probability). In the CV word condition, the stimulus /deI/ ("day") acted as the standard (87.5% probability), and the stimulus /geI/ ("gay") acted as the deviant (12.5% probability). A total of 300 sweeps were obtained to each deviant stimulus type. In all speech stimulus blocks, the stimuli were presented in a pseudorandom order, so that no fewer than two standards were presented between deviant stimuli. Subjects in Group A were presented with the speech stimuli at an SOA of 610 msec, and subjects in Group B were presented with the speech stimuli at an SOA of 900 msec. In addition, the deviant stimulus in each test condition was presented alone in a single block of 350 stimuli (the "deviant alone" condition).

Lastly, the tone and speech stimuli were also presented in a behavioral-discrimination condition to determine how easily the subjects were able to actively discriminate between the auditory stimuli. The behavioral condition for the simple tones consisted of one multiple deviant stimulus block (150–180 stimuli, including 25–30 deviant stimuli; SOA = 1500 msec). The behavioral conditions for the speech stimuli (nonwords and words) consisted of two single contrast blocks per test condition, with each stimulus acting as the standard in one block and as the deviant in the other block (100 stimuli per block, including 10 deviant stimuli; SOA = 1500 msec).

Procedure

Subjects were tested in an acoustically and electrically shielded room. During the ERP recordings, subjects were instructed to ignore the auditory stimuli and watch a silent, subtitled video of their choice. Throughout the experiment, stimuli were presented binaurally using the Neuroscan®, STIM system and EAR®, insert earphones. All tone and speech stimulus blocks were presented in a random order across subjects. The behavioral discrimination tasks were always administered following the passive discrimination condition to ensure that there was no carryover of attention that would affect the recorded automatic ERPs. During the behavioral discrimination tasks, subjects were instructed to actively listen to the sounds and press a button on the Stim response pad when they detected a deviant stimulus. Behavioral discrimination blocks were presented in a random order across subjects. The ERP measurements and behavioral tasks lasted between two and two and a half hours per subject, including breaks.

Electroencephalographic (EEG) Recording

A 32-channel ECI electrode cap was used to record the nose-referenced EEG (500 Hz sampling rate). All EEG measurements were made using a Neuroscan®, SynAmps amplifier and Scan 4.1 acquisition software. Recordings were obtained from the following tin electrodes (FP1, FP2, FPZ, FZ, FCZ, CZ, CPZ, PZ, F7, F3, F4, F8, FT7, FC3, FC4, FT8, T3, C3, C4, T4, TP7, CP3, CP4, TP8, T5, P3, P4, and T6), which were positioned according to the 10-20 International Electrode System (Jasper, 1958). The quality of the EEG recordings was monitored during data acquisition, and the continuous EEG data were stored on the computer for off-line processing. Electrodes attached to the right and left outer canthi monitored horizontal eye movements, and vertical eye movements were monitored by electrodes attached above and below the orbit of the left eye. Electrodes placed on the left and right mastoids (M1 and M2, respectively) were later used as a combined reference for the averaged data.

Data Analysis

The continuous raw data files were first subjected to ocular movement correction using a computational method available with Neuroscan®, SCAN software that incorporates regression analysis and artifact averaging to remove eye-blink artefacts (Semlitsch et al, 1986). The continuous data were divided into single sweeps of 700 msec, including a 100 msec pre-stimulus baseline. Individual epochs were subjected to a linear detrend over the entire epoch. This was followed by baseline correction based on the pre-stimulus interval. Artifact rejection (±100 µV), averaging, and digital bandpass filtering (1–30 Hz) were then carried out on the standard and deviant stimulus epochs in each block. The first 10 sweeps of each epoched file were excluded from the averaging process to reduce N1 amplitude variation associated with the start of the stimulation sequence (Pekkonen et al, 1995; Lavikainen et al, 1997; Sinkkonen and Tervaniemi, 2000). The same analyses were carried out on the "deviant alone" epochs. All of the processed average files were then individually re-referenced to the mastoids and baseline corrected. As the MMN inverts its polarity at electrodes below the level of the Sylvian fissure, arithmetic re-referencing of the contributing ERP waveforms to the average of the mastoids maximizes the MMN amplitudes at the frontal electrodes (Näätänen, 1995), thereby increasing the signal-to-noise ratio (Schröger, 1998).

Difference waveforms for all the stimulus contrasts were calculated by subtracting the averaged response to the deviant stimulus when it was presented alone, from the averaged response to the deviant stimulus when it was presented in the oddball paradigm. As mentioned in the introduction, this method of computing the difference waveform is to ensure that the mismatch response reflects only context-dependent differences between the standard and deviant stimuli. For comparison purposes, deviant minus standard difference waveforms were also calculated. Grand average waveforms for all the stimulus contrasts were calculated for each group of subjects and were re-referenced to the mastoids and baseline corrected before visual inspection and statistical analyses. Using the grand average data, the electrode with the maximum MMN, or where the responses "tended to be largest" (Paavilainen et al, 1991), was determined for each condition. The maximum electrode was then used to determine the peak latency of the MMN in the grand average data for each condition.

The grand average data obtained in response to the tone stimuli were pooled (i.e., n = 30) as the testing conditions and stimuli given were identical across the two groups. All subsequent group and individual measurements were made on the difference waveforms. The MMN peak latency was determined as the latency of the most negative point on the grand average difference waveform within a 100–220 msec time window after the onset of the deviant stimulus (Wunderlich and Cone-Wesson, 2001). For each individual average difference waveform, the MMN mean amplitude was measured over a 40 msec time window centered at this peak latency, for all scalp locations (Winkler et al, 1998). The latency of the MMN for each individual difference waveform was determined using this same 40 msec time window. When the P3a component was observed in the grand average data following the MMN response, the peak latency of the P3a response was measured from the central (CZ) grand average difference response as the latency of the most positive peak between 200–400 msec. The individual P3a mean amplitudes were measured over a 40 msec interval centered at the peak latency, and the individual P3a latencies were measured at the most positive peak within that same 40 msec time window (Winkler et al, 1998).

The use of mean amplitude measurements, rather than peak amplitude or area, was chosen based on recommendations by Picton et al (2000). Mean amplitudes are considered to be more stable than amplitudes at a fixed latency, and the careful use of a priori chosen time windows for mean amplitude measurements can be adjusted to include those parts of the difference waveform that are of interest, whether or not they contain clearly defined peaks. In addition, although this mean measurement can be easily converted to an area measurement by multiplying by the time period, any changes in scalp distribution of the ERP within the chosen time period can render the resultant measurements difficult to interpret (Picton et al, 2000).

Statistical Analysis

Two-tailed t-tests (dependent samples) were used to determine if the mean amplitudes of the MMN grand average responses and P3a grand average responses (i.e., the mean of individual amplitudes taken from individual grand average waveforms) significantly differed from zero for each condition. The grand average MMN and P3a responses were compared across stimulus conditions using one-way analysis of variance (ANOVA) with dependent measures for the tone conditions, and paired t-tests for the speech conditions. Group differences across stimulus conditions were compared using two-way ANOVAs. Greenhouse-Geisser corrections were applied when appropriate.

In the behavioral discrimination tasks, the hit rate (HR) was calculated by dividing the number of correct responses to targets by the number of target events. Responses given within a 100–1450 msec interval from the onset of the deviant stimulus were accepted, whereas responses falling outside this time interval were counted as "false alarms" (FA), which are defined as false positive behavioral responses. The FA rate was calculated by dividing the number of false positive responses by the total number of all responses. The behavioral results from each condition were compared using an ANOVA with dependent measures. Greenhouse-Geisser corrections were applied when appropriate. It was hypothesized that the simple tone stimulus contrasts and the speech stimulus contrasts would be easily discriminable to the subjects, as previously indicated by extensive pilot testing using these stimuli.

RESULTS

The grand average ERP waveforms at FCZ for the tone stimulus conditions are presented in Figure 1, and the grand average ERP waveforms at FCZ for the speech

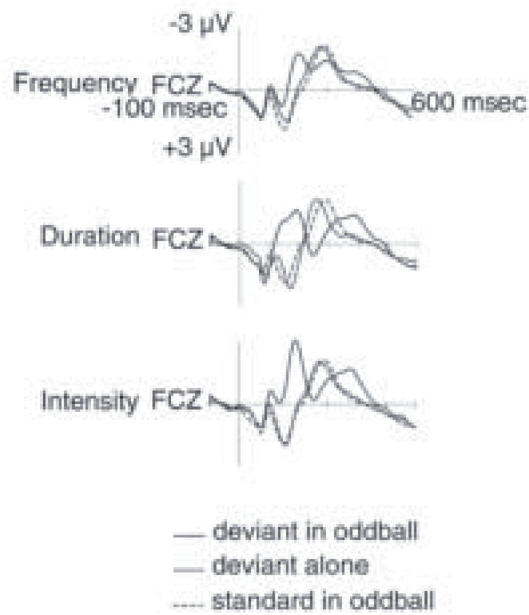

Figure 1. Re-referenced grand average ERPs at FCZ for the tone stimulus conditions (frequency, duration, and intensity) (n = 30).

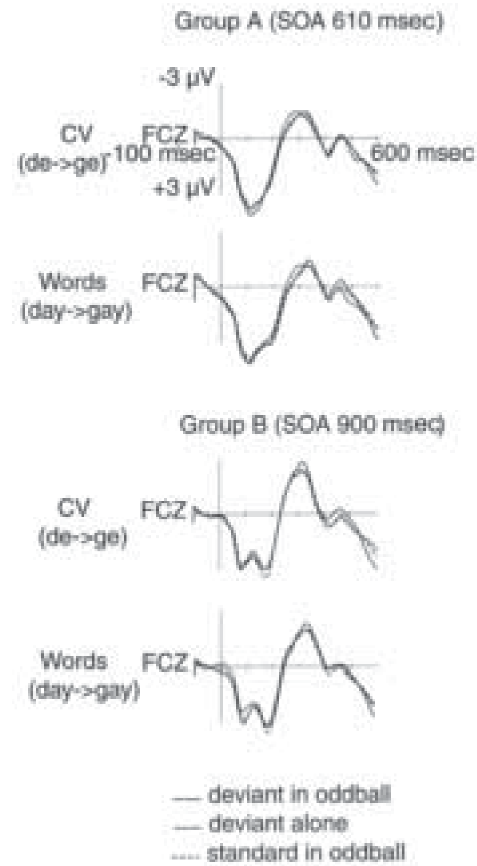

Figure 2. Re-referenced grand average ERPs for Group A and Group B at FCZ for the speech stimulus conditions (CV nonword and CV word conditions) (n = 15 each).

stimulus conditions are presented in Figure 2. The grand average ERP waveforms across all the fronto-central electrodes were of a similar pattern to those shown at FCZ.

MMN and P3a Responses to Simple Tones

The electrode at which the MMN response was largest in the grand average tone data was FCZ for all conditions, and the grand average difference waveforms at FCZ in response to the tone stimuli are shown in Figure 3. All the deviant stimulus types elicited visually well-defined MMN responses, which inverted their polarity at the mastoid electrodes, below the Sylvian fissure. The grand average mean amplitudes and peak latencies of the MMN components at FCZ for each tone condition are shown in Table 2.

Grand Average MMN Response

The mean amplitudes of the MMN response in each tone condition were significantly different from zero ($p < 0.001$, two-tailed t-test for dependent samples). There was a significant difference in mean amplitude [$F(2, 58) = 21.956$, $p < 0.001$] and

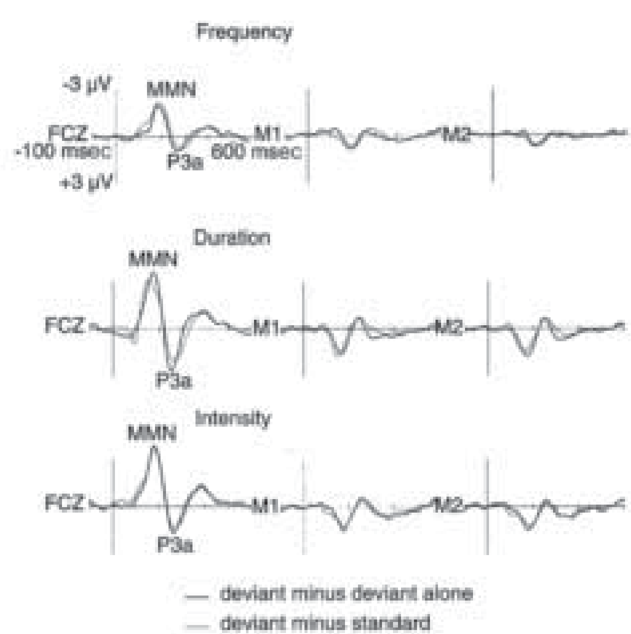

Figure 3. Re-referenced difference waves at FCZ (left column) for each tone condition (n = 30). The un-re-referenced grand average difference waves at the mastoid electrodes (M1 = left mastoid, M2 = right mastoid) indicate polarity reversal (middle and right columns).

Table 2. Grand Average Mean Amplitudes and Peak Latencies of MMN and P3a Responses (with Standard Deviations) for n = 30 at FCZ, for the Tone Conditions

MMN	Frequency	Duration	Intensity
Mean Amplitude (μV)	-1.73** ± 1.27	-3.10** ± 0.83	-3.30** ± 1.66
Peak Latency (msec)	181.40 ± 12.15	176.47 ± 10.68	183.07 ± 11.31
P3a			
Mean Amplitude (μV)	0.82* ± 1.47	2.23** ± 1.16	1.41** ± 1.56
Peak Latency (msec)	262.53 ± 14.19	253.47 ± 11.81	265.20 ± 13.49

**p < 0.001
*p < 0.01

peak latency ([$F(2, 58) = 3.783$, $p < 0.05$] across the three tone conditions for the pooled data set (n = 30).

A priori contrasts indicated that the mean amplitudes in the duration and intensity conditions were both significantly larger than the mean amplitude in the frequency condition [$F(1, 29) = 35.798$, $p < 0.001$; $F(1, 29) = 34.484$, $p < 0.001$, respectively]. The mean amplitudes in the duration and intensity conditions, however, did not significantly differ from each other ($p > 0.05$). In addition, a priori contrasts indicated that the peak latency of the grand average MMN response in the duration condition was significantly earlier than the peak latency of the intensity condition [$F(1, 29) = 7.764$, $p < 0.01$].

Grand Average P3a Response

Visual inspection of the grand average difference waves revealed the presence of a positive component immediately following the MMN, which was identified as a P3a response. The P3a mean amplitude values were significantly different from zero across all contrasts ($p < 0.01$, two-tailed t-test for dependent samples). There was a significant difference in mean amplitude [$F(2, 58) = 14.092$, $p < 0.001$] and in peak latency [$F(2, 58) = 11.413$, $p < 0.001$] across the three tone conditions for the pooled data set (n = 30).

A priori contrasts indicated that the mean amplitude of the grand average P3a responses in the duration condition was significantly larger than the P3a mean amplitudes in both the frequency condition [$F(1, 29) = 32.687$, $p < 0.001$] and the intensity condition [$F(1, 29) = 8.122$, $p < 0.01$]. Further a priori contrasts indicated that the peak latency of the P3a response in the duration condition was significantly earlier than the P3a response in the frequency and intensity conditions [$F(1, 29) = 13.469$, $p < 0.01$; $F(1, 29) = 27.791$, $p < 0.001$, respectively].

MMN Responses to Speech Stimuli

There was no clear maximum electrode for the grand average speech data for either the CV nonword or CV word condition. The analysis of the speech data was, therefore, based on measurements at FCZ, similarly to the tone data, and as used by Shtyrov and Pulvermüller (2002). The grand average difference waveforms in response to the speech stimuli at FCZ are shown in Figure 4, and the grand average mean amplitudes and peak latencies at FCZ for each speech condition are shown in Table 3.

Grand Average MMN Response

The MMN mean amplitude values for Group A and Group B's responses to the speech stimuli were not significantly different from zero for any contrast ($p > 0.05$, two-tailed t-test for dependent samples). Similar measurements of mean amplitude at CZ and PZ also did not yield significant results. Paired t-tests within groups revealed that for both Group A and Group B, there were no significant differences in mean amplitude across the two conditions (CV nonword and CV word), at either FCZ, CZ, or PZ.

Two-way ANOVAs indicated that there was no significant difference in MMN mean amplitude between the Groups for any condition, at any of the three electrodes (FCZ, CZ, and PZ). There was no apparent P3a response upon visual inspection of the waveform data; therefore, P3a measurements were not carried out.

Table 3. Grand Average MMN Mean Amplitudes and Peak Latencies (with Standard Deviations) for Group A (n = 15) and Group B (n = 15) at FCZ, for the Speech Conditions

	Group A (SOA 610 msec)		Group B (SOA 900 msec)	
	CV nonwords	CV words	CV nonwords	CV words
Mean Amplitude (µV)	0.16 ± 0.71	-0.13 ± 0.72	-0.08 ± 0.90	-0.00 ± 0.86
Peak Latency (msec)	162.53 ± 12.47	126.13 ± 16.06	195.73 ± 15.67	166.13 ± 13.97

Individual Data Analysis

An MMN response to the CV nonword and CV word stimuli could not be identified by visual inspection of the waveforms, at either a short or long SOA. Statistical analysis of the grand average data, as described above, confirmed this observation.

The individual variability of the MMN response amplitudes and latencies to the speech stimuli was noticeably higher than the individual variability of the simple tone MMN responses, as evidenced by the example in Figure 5. This high variability of individual speech MMNs may have substantially contributed to the poor grand average speech MMN results.

Although the individual MMN responses to the speech stimuli appear to be less consistent than the responses to the tone (e.g., duration) stimuli, it is likely that the poor grand average speech results may have been simply due to small individual MMN response amplitudes. Therefore, further investigation of the speech data at an individual level was carried out to determine whether the individual MMN responses to the

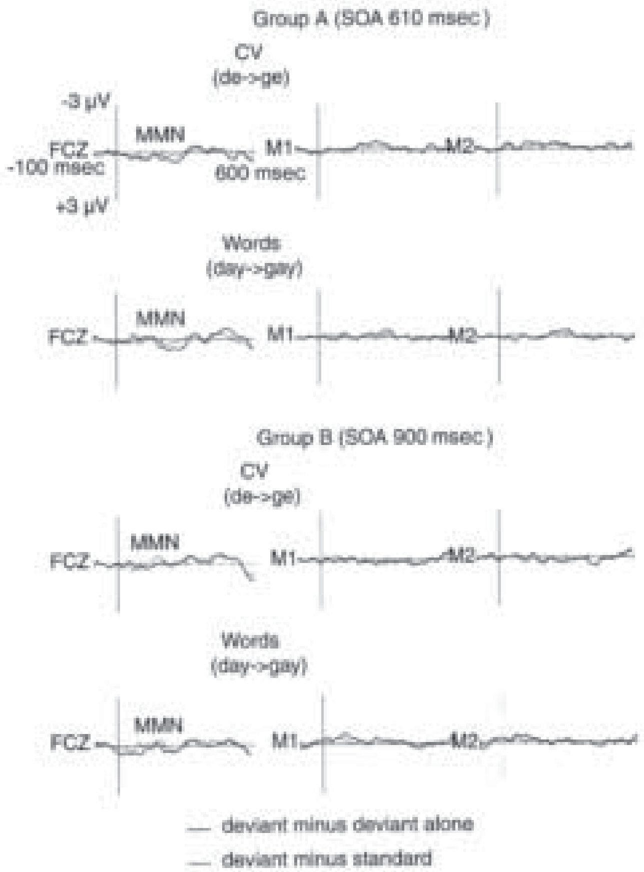

Figure 4. Re-referenced difference waves at FCZ (left column) for each speech condition (n = 15 for each Group). The un-re-referenced grand average difference waves at the mastoids (middle and right columns) indicate polarity reversal.

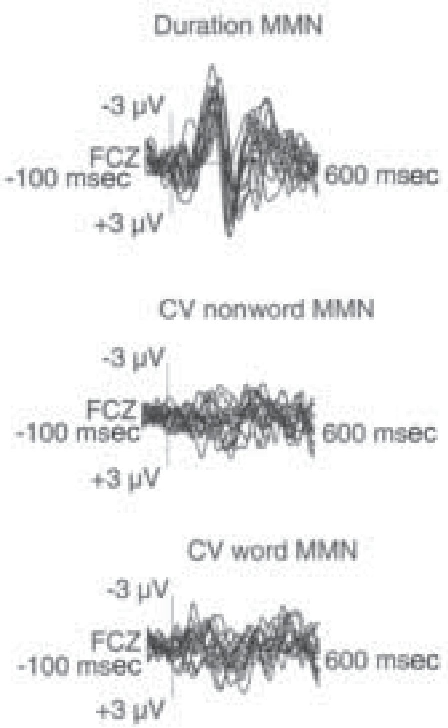

Figure 5. Superimposed individual re-referenced MMN waveforms at FCZ (Group A only, n = 15) for the tone duration condition (top), CV nonword condition (middle), and CV word condition (bottom).

Table 4. Percentages of Individual MMNs (n = 30) Present for Each Condition, at $p < 0.05$ and $p < 0.1$

Condition	$p < 0.05$	$p < 0.1$
Frequency	66.67%	76.67%
Duration	96.67%	96.67%
Intensity	96.67%	100%
CV nonword	10%	23.33%
CV word	10%	30%

speech stimuli were simply not distinguishable from noise.

The individual MMN responses in each condition were first identified as a relative negativity following the peak latency of the N1 response in the contributing ERP waveforms. Ten subaverages were generated using individual epoched data for the deviant waveform and deviant alone waveform responses for each subject, with the number of sweeps per subaverage calculated based on the number of accepted sweeps for that individual subject in each condition (ranging between 26–37 sweeps per subaverage for the tone conditions and 25–35 sweeps per subaverage for the speech conditions, across all subjects). The subaverages were each filtered, re-referenced to the mastoids, and baseline corrected. For each condition, the peak latency from the individual grand average MMN responses was used as an a priori selected time point for peak detection within each subaverage. Paired t-tests were then used to compare the amplitudes of the peaks in the deviant subaverages with the peaks in the deviant alone subaverages. The MMN response for a particular condition was deemed to be "present" if the difference between the average deviant and deviant alone peaks at the selected time point was significant.

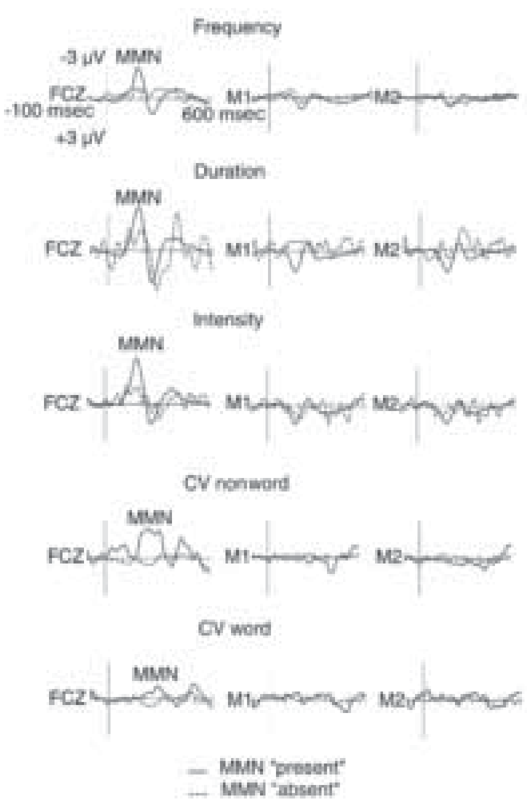

Figure 6. Grand average MMNs "present" vs. "absent" ($p < 0.05$) at FCZ for the frequency condition (n "present" = 20/30), duration condition (n "present" = 29/30), intensity condition (n "present" = 29/30), CV nonword condition (n "present" = 3/30) and the CV word condition (n "present" = 3/30) (left column). The un-re-referenced grand average difference waves at the mastoids indicate polarity reversal (middle and right columns).

Based on the above method of analysis, the percentages of individual MMNs present for each condition were calculated and are shown in Table 4. The grand average MMNs "present" and "absent" for each condition are shown in Figure 6.

These individual data analysis results provide further confirmation that the lack of MMN response to the speech stimuli evident in the grand average data reflected poor responses at the individual level, across almost all of the subjects.

Behavioral Data

The hit rates and false alarms across conditions for both Group A and Group B are shown in Table 5. The behavioral task results indicate that both the simple tone stimulus

Table 5. Hit Rate (HR) and False Alarm (FA) Mean Percentages and Standard Deviations for Each Behavioral Condition

		Group A		Group B	
		HR (%)	FA (%)	HR (%)	FA (%)
Tones	Frequency	96.43 ± 6.47	0.23 ± 0.89	98.67 ± 3.52	1.64 ± 2.84
	Duration	97.87 ± 6.07		95.87 ± 6.70	
	Intensity	100 ± 0		100 ± 0	
Speech	de ➔ ge	99.33 ± 2.58	0	89.33 ± 20.86	4.97 ± 12.47
	ge ➔ de	99.33 ± 2.58	0	98.00 ± 5.61	0.67 ± 2.35
	day ➔ gay	97.33 ± 5.936	0	99.33 ± 2.58	1.21 ± 3.20
	gay ➔ day	99.33 ± 2.58	0.61 ± 2.35	98.00 ± 4.14	0.67 ± 2.58

contrasts and the speech stimulus contrasts were easily discriminable to the subjects under attentional conditions.

DISCUSSION

The robustness and consistency of the MMN responses to the simple tone stimuli indicate that strong MMN responses can be obtained under the general testing conditions of the present study. However, the MMN responses to the CV syllable contrasts beginning with the voiced plosive (e.g., d/g) resulted in MMN waveforms that were small and difficult to distinguish from noise, in both the individual and group data from healthy adults. These results are consistent with the findings of Wunderlich and Cone-Wesson (2001), demonstrating that when carefully controlled methodological designs and strict methods of analysis are applied, robust MMN responses to CV syllable contrasts may be difficult to obtain.

MMN and P3a Responses to Simple Tones

The robustness of the MMN and P3a responses elicited by the simple tone stimuli was as to be expected, based on previous literature (Paavilainen et al, 1991). In particular, the P3a response commonly follows the MMN component (Lyytinen et al, 1992) and is proposed to reflect an automatic attention switch (Sams et al, 1985; Winkler et al, 1997). The duration change in this study elicited the largest MMN and P3a responses, at the earliest latencies, indicating that it may have elicited a stronger automatic switch to attention than the other deviant stimuli. A possible cause of this phenomenon is the well-established finding that a duration decrement of less than 200 msec causes a perception of both duration and loudness decrement (Hawkins and Presson, 1986).

Interestingly, the amplitudes of both the duration change and the intensity change MMNs were larger than the frequency change MMN, which indicates that the participating subjects found it more difficult to automatically detect frequency change rather than duration or intensity change. This finding may partly explain the lack of measurable MMN response in the same subjects to the fine acoustic contrasts embedded in the more complex speech stimuli, which were based entirely on formant (frequency) changes.

MMN Responses to Speech Stimuli

Individual Data Analysis

The individual data analyses indicate that despite the robust responses elicited by the simple tone stimuli in most of the subjects (66.67%–96.67%, $p < 0.05$), the MMN responses to the speech contrasts were either absent or of such small amplitude across the same subjects (10%, $p < 0.05$) that the group MMN data was not distinguishable from noise. In addition, the lack of clear mastoid reversal for the small group of subjects with MMNs “present” in the speech conditions (see Figure 6) suggests that the peaks in the speech-elicited difference waveforms may not in fact be true MMN responses but relative negativities. Given the reports in the literature that spectrally rich stimuli (such as speech) elicit larger MMNs than simple stimuli (Jaramillo et al, 1999; Tervaniemi et al, 2000), the difference in results between the simple tone and speech conditions in the present study are surprising yet consistent with the findings of Wunderlich and Cone-Wesson (2001). In particular, the poor MMN results in the speech stimuli conditions were also unexpected given the wealth of literature reporting robust MMN responses to CV syllable tokens contrasting the phonemes /d/ and /g/, both in children (Kraus et al, 1992; Kraus, McGee, Carrell et al, 1993; Kraus, McGee, Carrell, Sharma et al, 1995; Bradlow et al, 1999; Kraus et al, 1999; Uwer et al, 2002) and in adults (Kraus et al, 1992). Kraus et al (1992) obtained robust MMN responses in adults to a synthesized variant of the deviant /da/ with a synthesized variant of /ga/ as the standard. In contrast, however, studies by Dalebout and Stack (1999) and Dalebout and Fox (2000) investigated the MMN response in adults to variants of /da/ and /ga/ along a synthesized continuum and found it was not robust enough to be consistently identified in the individual response data. In particular, Dalebout and Fox (2000) obtained an MMN in response to /da/ versus /ga/ in only 40% of adult listeners.

All of the above-mentioned studies applied strict methods of data analysis to avoid confounding variables such as physical differences between the speech stimuli and to accurately identify individual MMN responses. However, it is not clear why Kraus et al (1992) were able to obtain robust MMN responses to the deviant /ga/ among /da/ standard stimuli, whereas the studies by Dalebout and Stack (1999) and Dalebout and Fox (2000) were not, when all three studies employed speech stimuli that were very similar in synthesis, duration, and intensity. Kraus et al (1992) obtained the robust responses in only a small sample size of adult subjects (n = 10), whereas the present study and the study by Dalebout and Fox (2000) were unable to obtain such robust responses in larger groups of adult listeners (n = 30). This observation is troublesome from the point of view that the MMN response to fine acoustic speech contrasts may not be consistent across the normal population.

Interestingly, the individual data analysis results indicated a slight increase in MMNs present for the CV word condition at a more lenient criterion level ($p < 0.1$), when compared with the CV nonword condition (23.33% present for CV nonwords, 30% present for CV words). Although there was no significant difference between the two conditions in the group data, these individual data results may indicate a trend toward the notion of linguistic meaningfulness facilitating the speech-elicited MMN response. This observation is consistent with the findings of several studies reporting larger MMN responses to linguistically meaningful stimuli (Pulvermüller et al, 2001; Shtyrov and Pulvermüller, 2002). In contrast, however, Wunderlich and Cone-Wesson (2001) reported only 25% of MMNs present for CVC words but 32% MMNs present for CV nonwords. These observations indicate that further investigation into the ability of linguistically meaningful stimuli to elicit the MMN in response to speech is needed.

It is reasonable to propose that the lack of robust MMNs elicited by the speech stimuli in the present study may be related to the notion that auditory sensory memory traces for plosive stops are weaker than for other phonemes, such as vowels, resulting in smaller mismatch responses (Diesch and Luce, 1997). However, the behavioral results of the present study indicate that all subjects were able to auditorily discriminate between the deviant and standard stimuli with a high degree of accuracy, suggesting that the d/g contrast used in the paradigm was not likely to have been too fine acoustically to be detected by the auditory system. Dalebout and Stack (1999) similarly reported a high percentage of correct identification (91%) of an easily discriminable /da/-/ga/ contrast during a behavioral task condition, while noting that only half of the subjects demonstrated valid MMN responses. The results of these studies indicate that the behavioral performance and automatic MMN responses of healthy adults do not necessarily correlate for fine acoustic speech contrasts, which indicates possible limitations to the reliability of the speech-elicited MMN in clinical situations.

Methodological Design

Similar to Wunderlich and Cone-Wesson (2001), great care was taken in choosing parameters commonly used in EEG recordings of this nature. For example, the

32-electrode montage adhered to the International 10-20 system (Jasper, 1958), and commonly used EEG recording parameters (e.g., sampling rate) were employed during data collection. The methodological design of the paradigm was based on commonly accepted oddball paradigms described in the MMN literature, including the probability of the deviant. The procedures employed during data analysis, and statistical analysis of the group and individual data, were chosen based on the recommendations made by several review papers (Lang et al, 1995; Schröger, 1998; Picton et al, 2000; Sinkkonen and Tervaniemi, 2000). In particular, the subtraction of the averaged deviant alone waveform response, rather than the averaged standard waveform response, from the average deviant waveform response, ensured that the difference waveform response contained no responses to physical differences between the stimuli. As demonstrated by Figures 3 and 4, visual inspection of the difference waveforms demonstrates that there can be distinct differences in MMN amplitude between the two types of subtraction waves, indicating the importance of using the most accurate method for computing the difference waveforms. Furthermore, the methods used in the individual data analysis were found to be the most accurate and effective method of identifying the individual MMN responses in the current data, only after extensive research into a variety of possible methods recently established in the literature. Furthermore, the fact that robust responses were successfully elicited from the same subjects using the simple tone stimuli validates the procedures used for collecting data, suggesting that the poor speech results were not attributable to data collection difficulties or inconsistencies.

Therefore, it is likely that the poor speech MMN results are not attributable to these general methodological aspects of the study design but are more likely due to speech-specific aspects of the methodological design. For example, the generation of the speech stimuli must be carefully considered. The pitfalls of using naturally spoken syllables to create speech stimuli in an MMN paradigm have already been discussed above. However, synthesized speech stimuli may be harder to discriminate than natural speech stimuli due to the smaller set of available acoustic cues for accurate acoustic-phonetic processing (Duffy and Pisoni, 1992). In contrast to the study by Wunderlich and Cone-Wesson (2001), several studies have reported robust MMN responses to speech stimuli with stop-consonant contrasts that were generated using synthesized speech (Kraus et al, 1992; Maiste et al, 1995; Kraus et al, 1999; Sharma and Dorman, 1999). The use of the SSG technique in the current study aimed to avoid the difficulties of acoustic differences between naturally spoken stimuli, without sacrificing as much "naturalness" as synthesized speech. However, the use of semisynthesized speech did not result in robust speech-elicited MMN responses.

Another speech-specific methodological aspect of the present study that must not be overlooked is the distractor task (viewing of a silent, subtitled video). It could be suggested that the use of a distractor task that involved language processing, albeit at a higher level of attention, may have contributed to the attenuation of the MMN response to the speech stimuli. Although it has been well established in the literature that the effectiveness of the distractor task is vital when attempting to obtain reliable MMN responses without the influence of attention, the possible effects of language-related distractor tasks (e.g., reading or watching videos with a soundtrack) on the automatic MMN responses has not yet been addressed thoroughly in the literature. Kraus, McGee, Carrell et al (1993) and Kraus et al (1992) successfully obtained MMN responses to speech stimuli in school-aged children and to adults by presenting the experimental auditory stimuli to one ear and allowing the subjects to watch a video and listen through the untested ear to the soundtrack (presented at an intensity level lower than the experimental stimuli). Interestingly, neither Dalebout and Fox (2000) nor Wunderlich and Cone-Wesson (2001) were able to obtain reliable MMN responses using this approach. However, the authors of these studies did not discuss the possible enhancing or attenuating effects of the language-related distractor task on the MMN results. In the current study, subjects were asked to watch silent videos that contained subtitles, while the stimuli were presented binaurally. The use of subtitled videos was considered appropriate given the length of the individual testing sessions (two to two and a half hours) and the necessity of preventing boredom (i.e., poor distraction) and/or fatigue, both of which may adversely affect the MMN results by either attenuating the amplitude of the MMN response or increasing the number of artifacts present in the EEG recording. To ensure that the language processing involved in reading the subtitles did not attenuate the MMN response to the experimental auditory speech stimuli, extensive post-hoc experiments were carried out to determine whether there was any difference in the MMN response related to the presence or absence of subtitles in the distractor task videos. The results of this testing indicated that not only was there no difference in the amplitude of the MMN response to either the tone or speech stimuli but the number of artifacts rejected during the recordings that used subtitled videos as a distractor was noticeably reduced compared with recordings that used videos without subtitles (and no sound). These results suggest that the poor MMN responses to the speech stimuli in the present study were not due to the reallocation of attentional resources or higher levels of language processing.

The SOAs used in the current study were based on SOAs commonly used in the MMN literature for tonal stimuli (610 msec) (Paavilainen et al, 1991) and speech stimuli (900 msec) (Shtyrov et al, 1998; Shtyrov, Kujala, Ilmoniemi et al, 2000; Shtyrov, Kujala, Lyytinen et al, 2000; Shtyrov and Pulvermüller, 2002). Lengthy SOAs in the order of 1–1.5 sec have also been used successfully with speech stimuli in previous studies (Kraus, McGee, Carrell et al, 1993; Maiste et al, 1995; Diesch and Luce, 1997; Näätänen et al, 1997; Kraus et al, 1999; Winkler, Kujala et al, 1999; Pulvermüller et al, 2001). Stimulus presentation rate can directly affect the length of a testing session, which is widely regarded as an important consideration in the design of MMN studies, as the MMN can be distorted by fatigue (Lang et al, 1995) and long-term habituation (McGee et al, 2001). Hence, the most effective SOA is the smallest at which a reliable MMN can be obtained in response to the deviant stimulus. The present study, however, demonstrated that neither a fast (610 msec, Group A) nor slow (900 msec, Group B) stimulus rate seemed to improve the responses to the speech stimuli.

Interestingly, higher percentages of individual MMNs present were noted at an SOA of 900 msec (13.33% for CV words, and for CV nonwords), rather than 610 msec (6.67% for CV words, and for CV nonwords) ($p < 0.05$). While it is reasonable to propose that the longer SOA would be expected to yield better results due to the length and complexity of the speech stimuli, it is important to consider the possibility that this increase in measurable MMN responses at 900 msec may also reflect the presence of refractory artifact. Walker et al (2001) proposed that the use of difference waves to identify the MMN, particularly the deviant alone extraction method, may be susceptible to artifacts related to neuronal refractoriness. It is well-known that increasing the interstimulus interval (ISI) between adjacent sounds in a stimulus sequence results in increased amplitudes of components such as N1 and P2. Therefore, subtraction of the deviant alone waveform (elicited using ISIs similar to the standard stimulus in the oddball paradigm) from the deviant-in-oddball waveform (much longer ISIs) may result in a negative baseline shift that reflects ISI-dependent amplitude changes in addition to, or instead of, deviance detection.

In their study, Walker et al (2001) reported no significant differences in amplitude for simple tone stimuli between the different extraction methods commonly used in MMN studies (deviant-minus-standard, deviant-minus-deviant alone, and reverse oddball procedures). However, the authors noted that increasing the ISI during the deviant alone condition, to more closely match the ISIs of the oddball deviant stimuli, greatly increased the N1-P2 amplitudes. Thus, activity represented in the difference waveform is very sensitive to changes in the ISI of the deviant in a deviant alone condition. However, Walker et al (2001) showed that when the ISI of the deviant stimuli in the deviant alone condition was the same as or similar to the ISI of the standard stimuli in the oddball sequence, the negative baseline shift in the difference waveform supported the existence of an MMN response. Therefore, it appears that the possibility of artifact contamination of the difference waveform is limited by presenting the deviant as "standard-like" as possible in the deviant alone condition (i.e., using the same ISI as the standard in the oddball sequence), as carried out by the present study. Therefore, it is unlikely that the increase in measurable

MMNs at a longer SOA observed in the current study is due to neural refractory artifact. Nevertheless, while the use of the deviant alone subtraction method remains justified by the fact that it is more time efficient then reverse oddball procedures and more accurate than deviant-minus-standard procedures, its susceptibility to neural refractory artifacts must still be carefully monitored in all MMN studies.

Habituation and Adaptation

According to Sallinen and Lyytinen (1997), significant attenuation of the MMN response can also result from adaptation of the neural processes that generate the MMN response over lengthy recording sessions. McGee et al (2001) found that the MMN in young adults, school-aged children, and guinea pigs elicited by synthesized speech syllable tokens (including /da/-/ga/) can experience significant habituation over time, adversely affecting the signal-to-noise ratio (SNR). The authors reported that even in the absence of subject fatigue, the MMN responses to the speech stimuli declined in magnitude (habituated) after 10–15 minutes of testing. However, the MMN response was significantly restored following rest breaks. In order to avoid the possible long-term habituation of the MMN response to the speech stimuli in the current study, the experimental blocks of speech stimulus trials were presented in a random order for each subject and were randomly interspersed with blocks of simple tone stimulus trials (essentially ensuring that there were several "breaks" for the speech-evoked MMN response). In addition, the subjects remained attentive to the videos throughout the MMN recording session and experienced regular rest breaks. Under these testing conditions, it is less likely that habituation of the MMN response could have resulted in such overwhelming attenuation of the MMN amplitude.

Further to the issues of neural adaptation and habituation of the MMN response, Wunderlich and Cone-Wesson (2001) suggested that a possible explanation for their poor speech MMN results may lie within the theory of adaptation and lateral inhibition (May et al, 1999). It has been proposed that the frequently repeated "standard" stimuli in an oddball presentation sequence develop a memory trace that may represent features of the auditory stimulus (e.g., duration, frequency). The mechanism that generates the MMN is activated when a "deviant" stimulus is presented while the standard memory trace is still active (Näätänen and Winkler, 1999). May et al (1999) proposed that the MMN response may be attributed to neuronal adaptation and lateral inhibition, in that for simple tones presented in an oddball sequence, a specific pattern of adaptation and lateral inhibition is generated across parts of the auditory cortex. The MMN response to the deviant tone is generated by a neuronal population that responds to the deviant stimulus while under the inhibitory influence of the neuronal activity evoked by the standard stimulus. Wunderlich and Cone-Wesson (2001) proposed that for spectrally complex stimuli (such as speech), repetitive presentation of the standard stimulus may generate a broad pattern of adaptation and lateral inhibition that may affect most of the cortical areas subsequently evoked by the deviant stimulus. Given the fine (yet clearly discriminable) acoustic differences between the standard and deviant stimuli in both the current study and the study by Wunderlich and Cone-Wesson (2001), it is reasonable to propose that the amount of overlap in the cortical areas activated by the deviant and standard stimuli may be of sufficient magnitude that the resultant pre-attentive MMN responses to the deviant stimuli are too small to be accurately detected and recorded by the scalp electrodes. Furthermore, the effects of adaptation and lateral inhibition may in fact hinder the recovery of the response from any habituation that has occurred over time (Rosburg et al, 2002). That is, any overlap in groups of neuronal populations processing the standard and deviant stimuli may reduce the amount of response recovery from habituation (Rosburg et al, 2002).

CONCLUSION

The results of the present study indicate that when confounding variables such as acoustic/physical differences between the deviant and standard stimuli are appropriately controlled for by the methodological design of the study, MMN responses to fine-grained CV speech stimulus contrasts are not always robust. The results of this study support the findings of Wunderlich and Cone-Wesson (2001) and suggest that further investigations using CV and CVC contrasts are needed to determine the optimal speech stimulus contrasts for the MMN to accurately reflect pre-attentive language processing at the word level. These further studies will require careful consideration of methodological design and data analysis procedures, in order to avoid the influence of confounding variables that may provide false MMN responses. Furthermore, the possible effects of habituation, adaptation, and lateral inhibition on the neuronal populations underlying the MMN response should be determined, in order to facilitate the establishment of appropriate MMN stimulus paradigms for the investigation of speech processing at the word level in normal and pathological populations.

REFERENCES

Aaltonen OJ, Tuomainen J, Laine M, Niemi, P. (1993). Cortical differences in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN). *Brain Lang* 44:139–152.

Alku P, Tiitinen H, Näätänen R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clin Neurophysiol* 110:1329–1333.

Aulanko R, Hari R, Lounasmaa OV, Näätänen R, Sams M. (1993). Phonetic invariance in the human auditory cortex. *Neuroreport* 4:1356–1358.

Bradlow AR, Kraus N, Nicol TG, McGee TJ, Cunningham J, Zecker SG. (1999). Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *J Acoust Soc Am* 106:2086–2096.

Ceponiene R, Service E, Kurjenluoma S, Cheour M, Näätänen R. (1999). Children's performance on pseudoword repetition depends on auditory trace quality: evidence from event-related potentials. *Dev Psychol* 35:709–720.

Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Näätänen R. (1998). Development of language-specific phoneme representations in the infant brain. *Nat Neurosci* 1:351–353.

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

Dalebout SD, Stack JW. (1999). Mismatch negativity to acoustic differences not differentiated behaviourally. *J Am Acad Audiol* 10:388–399.

Deacon D, Nousak JM, Pilotti M, Ritter W, Yang C-M. (1998). Automatic change detection: does the auditory system use representations of individual stimulus features or gestalts? *Psychophysiology* 35:413–419.

Dehaene-Lambertz G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport* 8:919–924.

Diesch E, Luce T. (1997). Magnetic mismatch fields elicited by vowels and consonants. *Exp Brain Res* 116:139–152.

Duffy SA, Pisoni DB. (1992). Comprehension of synthetic speech produced by rule: a review and theoretical interpretation. *Lang Speech* 35:351–389.

Giard M, Perrin F, Pernier J, Bouchet P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* 27:627–640.

Hawkins HL, Presson JC. (1986). Auditory information processing. In: Boff KR, Kaufman L, Thomas JP, eds. *Handbook of Perception and Human Perception*. New York: Wiley, 26.1–64.

Hillyard SA. (2000). Electrical and magnetic brain recordings: contributions to cognitive neuroscience. In: Gazzaniga MS, ed. *Cognitive Neuroscience: A Reader.* Oxford: Blackwell Publishers, 25–37.

Jaramillo M, Alku P, Paavilainen P. (1999). An event-related potential (ERP) study of duration changes in speech and non-speech sounds. *Neuroreport* 10:3301–3305.

Jasper HH. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol* 10:371–375.

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

Kent RD, Read C. (1992). *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group.

Korpilahti P, Krause CM, Holopainen I, Lang AH. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain Lang* 76:332–339.

Kraus N, Koch DB, McGee TJ, Nicol TG, Cunningham J. (1999). Speech-sound discrimination in school-age children: psychophysical and neurophysiologic measures. *J Speech Lang Hear Res* 42:1042–1060.

Kraus N, McGee T, Carrell TD, Sharma A. (1995). Neurophysiologic bases of speech discrimination. *Ear Hear* 16:19–37.

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

Kraus N, McGee T, Carrell TD, Sharma A, Nicol T. (1995). Mismatch negativity to speech stimuli in school-age children. *Electroencephalogr Clin Neurophysiol Suppl* 44:211–217.

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

Kraus N, McGee T, Micco A, Sharma A, Carrell TD, Nicol T. (1993). 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.

Kutas M. (2000). Current thinking on language structures. In: Gazzaniga MS, ed. *Cognitive Neuroscience: A Reader*. Oxford: Blackwell Publishers, 431–446.

Kutas M, Van Petten C. (1994). Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher MA, ed. *Handbook of Psycholinguistics.* San Diego: Academic Press, 88–143.

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

Lavikainen J, Tiitinen H, May P, Näätänen R. (1997). Binaural interaction in the human brain can be noninvasively accessed with long-latency event-related potentials. *Neurosci Lett* 222:37–40.

Lyytinen H, Blomberg AP, Näätänen R. (1992). Event-related potentials and autonomic responses to a change in unattended auditory stimuli. *Psychophysiology* 29:523–534.

Maiste AC, Wiens AS, Hunt MJ, Scherg M, Picton T. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear Hear* 16:68–90.

May P, Tiitinen H, Ilmoniemi RJ, Nyman G, Taylor JG, Näätänen R. (1999). Frequency change detection in human auditory cortex. *J Comput Neurosci* 6:99–120.

McGee T, King C, Tremblay K, Nicol TG, Cunningham J, Kraus N. (2001). Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology* 38:653–658.

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. (1992). *Attention and Brain Function.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Näätänen R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38:1–21.

Näätänen R, Alho K. (1997). Higher-order processes in auditory-change detection. *Trends Cogn Sci* 1:44–45.

Näätänen R, Escera C. (2000). Mismatch negativity: clinical and other applications. *Audiol Neurootol* 5:105–110.

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

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.

Näätänen R, Paavilainen P, Alho K, Reinikainen K, Sams M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neurosci Lett* 98:217–221.

Näätänen R, Paavilainen P, Reinikainen K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neurosci Lett* 107:347–352.

Näätänen R, Paavilainen P, Titinen H, Jiang D, Alho K. (1993). Attention and mismatch negativity. *Psychophysiology* 30:436–450.

Näätänen R, Winkler I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 125:826–859.

Paavilainen P, Alho K, Reinikainen K, Sams M, Näätänen R. (1991). Right hemisphere dominance of different mismatch negativities. *Electroencephalogr Clin Neurophysiol* 78:466–479.

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

Picton T, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R, Jr, Miller GA, Ritter W, Ruchkin DS, Rugg MD, Taylor MJ. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37:127–152.

Pulvermüller F, Kujala T, Shtyrov Y, Simola J, Tiitinen H, Alku P, Alho K, Martinkauppi S, Ilmoniemi RJ, Näätänen R. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage* 14:607–616.

Rosburg T, Haueisen J, Sauer H. (2002). Habituation of the auditory evoked field component N100m and its dependence on stimulus duration. *Clin Neurophysiol* 113:421–428.

Sallinen M, Lyytinen H. (1997). Mismatch negativity during objective and subjective sleepiness. *Psychophysiology* 34:694–702.

Sams M, Aulanko R, Aaltonen O, Näätänen R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *J Cogn 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.

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

Semlitsch HV, Anderer P, Schuster P, Presslich O. (1986). A solution for reliable and valid reduction of ocular artifacts applied to the P300 ERP. *Psychophysiology* 23:695–703.

Sharma A, Dorman MF. (1999). Cortical auditory evoked potential correlates of categorical speech perception of voice-onset time. *J Acoust Soc Am* 106:1078–1083.

Sharma A, Dorman MF. (2000). Neurophysiologic correlates of cross-language phonetic perception. *J Acoust Soc Am* 107:2697–2703.

Shtyrov Y, Kujala T, Ahveninen J, Tervaniemi M, Alku P, Ilmoniemi RJ, Näätänen R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. *Neurosci Lett* 251:141–144.

Shtyrov Y, Kujala T, Ilmoniemi R, Näätänen R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *Neuroimage* 12:657–663.

Shtyrov Y, Kujala T, Lyytinen H, Kujala J, Ilmoniemi R, Näätänen R. (2000). Lateralization of speech processing in the brain as indicated by mismatch negativity and dichotic listening. *Brain Cogn* 43:392–398.

Shtyrov Y, Pulvermüller F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport* 13:521–525.

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

Tallal P, Miller SL, Bedi G, Byma G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM. (1996). Language comprehension in language-impaired children improved with acoustically modified speech. *Science* 271:81–84.

Tervaniemi M, Schröger E, Sahe M, Näätänen R. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans—a mismatch negativity study. *Neurosci Lett* 290:66–70.

Titova N, Näätänen R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci Lett* 308:63–65.

Uwer R, Albrecht R, von Suchodoletz W. (2002). Automatic processing of tones and speech stimuli in children with specific language impairment. *Dev Med Child Neurol* 44:527–532.

Walker LJ, Carpenter M, Downs CR, Cranford JL, Stuart A, Pravica D. (2001). Possible neuronal refractory or recovery artifacts associated with recording the mismatch negativity response. *J Am Acad Audiol* 12:348–356.

Winkler I, Kujala T, Tiitinen H, Sivonen P, Alku P, Lehtokoski A, Czigler I, Csépe V, Ilmoniemi RJ, Näätänen R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36:638–642.

Winkler I, Lehtokoski A, Alku P, Vainio M, Czigler I, Csépe V, Aaltonen O, Raimo I, Alho K, Lang H, Iivonen A, Näätänen R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn Brain Res* 7:357–369.

Winkler I, Tervaniemi M, Näätänen R. (1997). Two separate codes for missing-fundamental pitch in the human auditory cortex. *J Acoust Soc Am* 102:1072–1082.

Winkler I, Tervaniemi M, Schröger E, Wolff C, Näätänen R. (1998). Preattentive processing of auditory spatial information in humans. *Neurosci Lett* 242:49–52.

Wunderlich JL, Cone–Wesson BK. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *J Acoust Soc Am* 109:1526–1537.