Mismatch negativity reflects sensory and phonetic speech processing

Marc F. Joanisse[a], Erin K. Robertson[b] and Randy Lynn Newman[c]

Correspondence to Dr Marc F. Joanisse, PhD, Department of Psychology, The University of Western Ontario, London, ON N6A 5C2, Canada
Tel: +1 519 661 2111 (ext) 86582; fax: +1 519 661 3961; e-mail: marcj@uwo.ca

We examined phonetic and sensory processes in speech perception using mismatch negativity, an event-related potential component congruent with discrimination, but which occurs for unattended stimuli. Adult listeners (*N*=16) heard a repeated standard (the syllable 'da') that was interrupted infrequently by a phonetically different 'deviant' syllable ('ba'). The acoustic difference between standard and deviant was manipulated to create both acoustically Strong and Weak deviant stimuli. Mismatch negativities in response to the Strong deviant were significantly greater than those for the Weak deviant, in spite of the fact that both represented stable instances of the phonetic category. The data suggest that the mismatch negativity component can be strongly influenced by sensory factors beyond what is predicted by overt categorization and discrimination judgments. *NeuroReport* 18:901–905 © 2007 Lippincott Williams & Wilkins.

Introduction

A central issue in speech perception research concerns how humans translate auditory sensory information into a phonetic representation, and especially whether speech-specialized mechanisms are used for this purpose [1]. Much of the evidence pertaining to this debate comes from the finding that many speech sounds are categorically perceived: that is, individuals tend to be sensitive to acoustic differences that differentiate phonetic categories, but ignore acoustic differences that are not phonetically relevant [2]. Recent event-related potential (ERP) studies have used the mismatch negativity (MMN) paradigm to examine neural mechanisms of categorical speech perception [3–6]. In an MMN paradigm, a repeated sound (the 'standard') is occasionally interrupted by an oddball stimulus (a 'deviant'). Subtracting standard from deviant ERPs reveals a negative-going waveform 100–300 ms poststimulus onset. It is suggested that the repeated standard activates a preexisting template in the auditory system, and that an MMN occurs when a subsequent sound violates this template [7].

A number of studies have observed that in addition to being sensitive to simple physical stimulus characteristics (e.g. frequency, duration and amplitude, [8,9]), MMNs are also observed in response to changes in complex temporal aspects of sounds such as duration, rhythm or temporal order [10,11]. This suggests the MMN technique is appropriate for studying acoustically complex stimuli such as speech. Accordingly, studies have examined mismatch responses to speech stimuli and have found MMNs tend to be increased in response to a between-category stimulus shift compared with a phonetically irrelevant within-category stimulus shift [5,6]. Such findings have been proposed to indicate that humans maintain domain-specific phonetic templates in auditory memory [12]. On the other hand, other studies have found MMNs for nonphonetic speech contrasts [8,13–15], raising the possibility that MMN does not strictly reflect phonetic processing and can also occur because of basic sensory mechanisms.

From a theoretical perspective, these findings reflect a growing debate about the role of sensory factors in speech processing. Behavioral studies have classically found non-linear response functions for speech sounds, leading to the generalization that listeners are equally sensitive to all between-category contrasts and ignore within-category contrasts [16]. Recent findings, however, have revealed more graded effects such that listeners respond more rapidly and accurately to prototype than nonprototype exemplars on a variety of tasks [17–19]. In this study, we examined whether similar effects might occur with MMNs. ERPs were recorded as participants heard a repeated syllable, interspersed with deviants from a different phonetic category. We manipulated the degree to which deviants differed from the standard to examine whether equal magnitude MMNs are observed for Strong and Weak acoustic contrasts.

Methods
Participants
A total of 16 neurologically healthy right-handed adults participated in this study (11 women, 5 men; mean age 25;4 years). All were native English speakers recruited from the University of Western Ontario community. Informed consent was obtained before the experiment commenced.

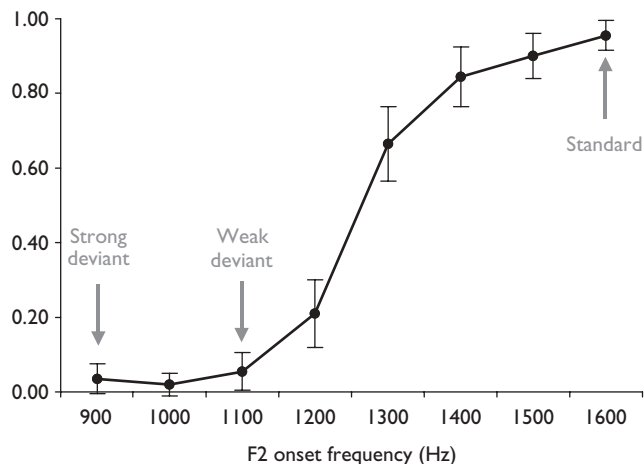All procedures were approved by the University of Western Ontario Office of Research Ethics.

Stimuli

Stimuli consisted of the syllables ba and da synthesized using a digital implementation of the Klatt parallel synthesizer [20]. A continuum of eight consonant–vowel syllables was generated by manipulating the onset frequency of the second formant (F2) transition from 900 to 1600 Hz, in 100 Hz steps (Table 1). From this continuum, we selected the da endpoint as the Standard, the ba endpoint item as the Strong deviant and a ba midpoint item as the Weak deviant (Fig. 1). Pilot data from 10 adults who did not participate in the ERP experiment indicated very similar categorization rates for both deviant items [$t(9)=0.802$, $P=0.44$]. Both deviant stimuli were also discriminated from the Standard at equally high levels of accuracy [Strong vs. Standard: M: 94%, (SE: 9%); Weak vs. Standard M: 97% (SD: 3%); $t(9)=1.5$, $P=0.28$]. Thus, although the Strong and Weak deviants differed in their acoustic distance to the repeated standard, both represented perceptually stable exemplars of the 'b' category.

Electroencephalogram recording

An MMN procedure was used [4,7] in which standards and deviants were randomly presented at an 8:1:1 ratio, with no fewer than two standards separating each deviant. Stimulus SOA was 800 ms. Audio was presented via an insert earphone to the right ear, and an earplug was placed in the left ear to attenuate ambient noise. Participants did not attend to the sounds during recordings. To help, a feature film was played during the session, without audio but with subtitles on.

Signals were recorded to disk at 500 Hz using a Synamps amplifier. Data were recorded from 64 Ag/AgCl sintered electrodes, mounted in the international 10–20 system using a placement cap and referenced to nose tip. Impedances were kept below 10 kΩ. Signals were filtered online at 60 Hz, and offline with a 0.1–20 Hz digital filter.

Voltages were averaged for each condition from −100 to 598 ms, baseline-corrected to the prestimulus interval. Trials with voltages ±75 μV were rejected before averaging. MMN components were quantified subjectwise using an automatic procedure that identified the amplitude of negative voltage peaks within two time windows (early: 80–160 ms; late: 180–280 ms), at each electrode. Peaks were averaged into seven scalp regions (midline frontal: Fz, FCz, Cz; left frontal: F1, FC1, F3, FC3, F5, FC5; right frontal: F2, FC2, F4, FC4, F6, FC6; midline postcentral: Pz, CPz, POz; left postcentral:

Table 1   Parameters used to synthesize ba–da items

Time (ms)	AV	F0	FI	F2	F3	F4
0	76	120	200		2550	3300
15	76	—	450	—	2550	3300
30	72	—	600	990	2600	3300
120	72	110	600	990	2600	3300
220	72	100	600	990	2600	3300
240	0	100	600	990	2600	3300

Parameters held constant: A1: 60; A2: 56; A3: 52; A4: 48; AB: 36; BGP: 100; BI: 50; B2: 70; B3: 100; B4: 250.
Dashes denote interpolated values.

Fig. 1   Categorization rates for the speech stimuli, reflecting categorical perception effects. Arrows indicate stimuli that were subsequently used as standard and deviants in the ERP study. ERP, event-related potential.

P1, P3, P5, CP1, CP3, CP5; right postcentral: P2, P4, P6, CP2, CP4, CP6; lateral: P7, P8, TP7, TP8, T7, T8), capturing the generalization that MMN is typically stronger at fronto-central than at postcentral and lateral sites.

Results

Average waveforms are plotted in Fig. 2. Two-way repeated measures analyses of variance (ANOVAs) were performed for the effects of stimulus condition (deviant vs. standard) and scalp region on voltage peaks. A separate ANOVA was run for each deviant type at the two time windows. A significant MMN should be marked by a stimulus × region interaction, in which differences between the standard and deviant are greater at frontal regions than at posterior and lateral regions. Follow-up analyses used paired t-tests to compare the standard and deviant at each site (one-tailed, reflecting the prediction of greater negativity values for deviants than standards).

The Strong deviant condition showed clear MMN effects at both time windows. Early window: there was a significant effect of stimulus condition [$F(1,15)=10.029$ $P<0.01$], no effect of electrode site [$F(6,90)=1.256$ $P>0.05$], and a stimulus × site interaction [$F(6,90)=3.035$, $P<0.01$]. Post-hoc tests revealed significant differences between standard and deviant at the midline, left and right frontal sites [$t(15)=3.91$, 3.03 and 4.46 respectively, $P<0.001$], but not at any of the posterior or lateral sites. Late time window: significant effects of stimulus condition [$F(1,15)=9.887$ $P<0.01$] and electrode site [$F(6,90)=33.289$ $P<0.001$] and stimulus × site interaction [$F(6,90)=2.623$ $P<0.01$]. Post-hocs showed significant differences for the midline and right frontal sites [$t(15)=3.29$ and 3.89, $P<0.001$] and midline and right posterior sites [$t(15)=2.95$ and 3.06, $P<0.001$].
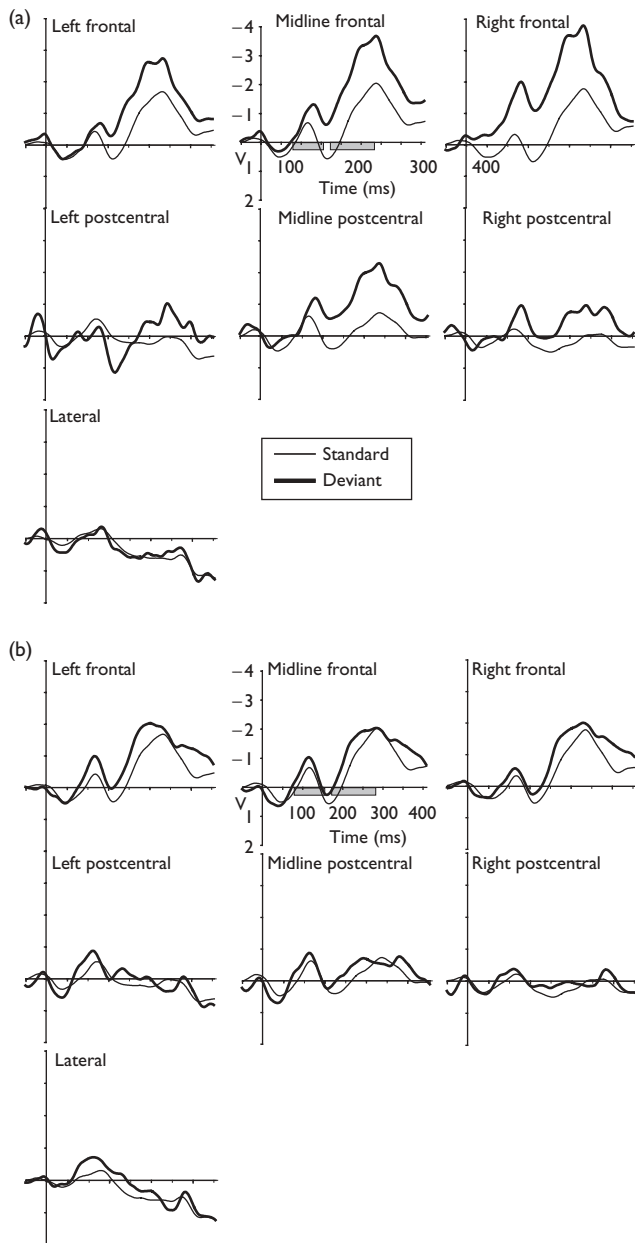
The weak deviant condition yielded less pronounced effects. Early time window: there was a significant effect of stimulus condition [$F(1,15)=5.907$ $P<0.05$] but no effect of electrode site and no interaction [$F(6,90)=0.976$ $P=0.45$ and $F(6,90)=0.745$ $P=0.62$, respectively]. [Although post-hoc tests are not indicated here, we observed significantly greater negativity for the weak deviant at the left frontal
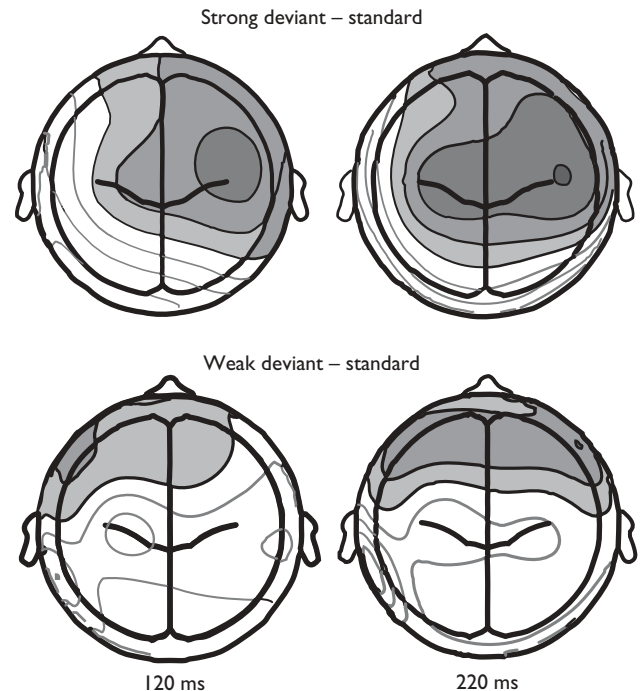
Fig. 2 Average waveforms for standard vs. deviant conditions. The strong deviant (a) yielded mismatch effects at both the early and late intervals (grey bars). The Weak deviant (b) showed an appreciably smaller effect.

Fig. 3 Scalp voltage maps (0.2 µV/line) of deviant minus standard subtraction. Shaded areas indicate negative voltages.

site, $t(15)=2.64$, $P<0.05$]. Late time window: we observed a main effect of electrode site [$F(6,90)=33.775$ $P<0.001$] but only a marginal effect of stimulus condition [$F(1,15)=3.954$, $P=0.07$] and no interaction [$F(6,90)=0.960$, $P=0.46$].

Strong and weak MMN waveforms were computed by subtracting the standard from deviants (Fig. 3). Peak negativities were calculated as above and submitted to two-way ANOVAs for the effects of deviant type (strong vs. weak) and electrode site. At the early time window there was a main effect of electrode site [$F(6,84)=3.787$ $P<0.005$], no effect of stimulus type [$F(1,14)=0.065$ $P<0.802$] and a significant interaction [$F(6,84)=2.755$ $P<0.025$]. The interaction confirms the observation that the Strong Deviant condition yielded mismatch effects across a greater dis-

tribution of sites. At the later time window, there was again an effect of electrode position [$F(6,84)=0.114$ $P<0.01$] but no effect of deviant type and no interaction [$F(1,14)=0.114$ $P=0.741$ and F(6,84) $P=0.646$]. Paired t-tests revealed significantly greater negativity for the strong deviant condition at the Right Frontal site [$t(15)=1.83$, $P<0.05$], although other differences were not significant.

Discussion

A key finding in speech research concerns the observation that humans categorically perceive many speech contrasts. Several prior studies have indicated MMNs also respect the categorical nature of speech contrasts. In the present experiment, we examined whether sensory factors also influence these effects. In our study, participants heard two deviant types; both were phonetically different from the standard; however, one was an acoustically prototypical category member, whereas the other was not. Behavioral studies suggest that listeners tend to identify both prototype and nonprototype exemplars as category members with equal proficiency. Our own data (Fig. 1) support this generalization. In contrast, the ERP data reported here indicate subtle differences in how prototypical and nonprototypical speech exemplars are processed preattentively. A prototypical phonetic distinction yielded much more reliable mismatch effects than the nonprototype, such that the Strong deviant showed significant MMNs at frontal electrode sites, whereas the Weak deviant showed less pronounced effects, especially at a later time window. Moreover, the scalp distribution of mismatch effects generated by the Weak deviant condition tended to be more limited.

These data seem consistent with prior studies of speech processing that have found subtle differences in how

listeners process prototype vs. nonprototype speech sounds, especially in on-line tasks [18,19]. These results indicate that listeners can show graded effects in phonetic processing owing to acoustic factors, contrary to what is typically predicted by the modular view that speech processing occurs separately from other auditory sensory function [1]. It is also interesting to note that listeners identified and discriminated both deviant types equally well when offline behavioral tests were used; participants ignored phonetically irrelevant low-level sensory information as they made overt judgments about the phonetic form of speech. This suggests that behavioral response measures tell only part of the story of how individuals translate an acoustic signal into a phonetic code.

From a methodological perspective, our data indicate that basic sensory mechanisms can influence the magnitude of MMN effects in response to speech contrasts. The auditory system appears to be maintaining the acoustic form of the standard in memory, such that mismatch effects are modulated by the degree to which the deviant physically differs from it. Note that this does not preclude the possibility that a phonetic mode is also being engaged by this paradigm; for instance, there is good evidence that phonetic factors are also at play in MMN effects, given that some (but not all) studies have found stronger negativity for categorical contrasts than for noncategorical contrasts. It, however, does indicate that sensory features can also play a role in these results.

An alternative explanation of the finding of greater MMNs for the Strong deviant is that the large difference between it and the standard led to a perceptual pop-out effect. That is, participants' attention might have been unduly aroused by the Strong deviant, but not by the Weak deviant. Such an attentional effect would have yielded an increased N2b component, which we might have mistakenly classified as an MMN [21]. We argue this cannot explain the present results, however. First, the negative-going deflection for the deviant stimuli began at around 80 ms poststimulus onset, appreciably earlier than what is expected for an N2b. Second, the Strong and Weak MMNs actually showed less difference at the 'late' period (around 250 ms post stimulus onset), the point at which an attentional pop-out would have manifested itself. Third, we note a tendency toward an inversion of the MMN at lateral sites, which again is consistent with an MMN rather than an N2b. In summary, the effects do not appear to be attention-driven.

The time course of mismatch negativity

We considered two time windows with respect to MMN effects, on the basis of earlier observations of MMNs at both relatively early points (10–120 ms; Refs [5,22]) and as late as 300 ms [8,15]. One possibility is that early and late MMNs reflect different levels of processing; for instance, temporal features could yield later MMNs than acoustically simpler cues such as frequency change [23]. A related hypothesis is that sensory processing and phonetic processing are reflected in early and late time points, respectively [24]. Neither of these seems to explain the present data, however. The key stimulus shift in this study had to do with the direction and rate of change of the F2 formant transition, which is arguably a rapid temporal acoustic cue. We, however, observed significant negativity at both the early and late time window, suggesting that early-going MMNs

do occur in response to rapid temporal features. Likewise, the idea that MMN time course reflects a sensory/phonetic distinction would predict stronger MMNs earlier on for the Strong deviant condition, but similar late MMNs for the Weak and Strong contrasts. Instead we found the Weak deviant condition showed less pronounced negativities at both the early and late time windows.

Conclusion

We observed greater MMN effects for a large acoustic shift signaling a between-category phoneme contrast, compared with an acoustically smaller but still phonetically relevant contrast. The data suggest that sensory factors do contribute to low-level phonetic processing, but that such factors are more easily observed in electrophysiology than in off-line behavioral measures.

References

1. Studdert-Kennedy M. On the dissociation of auditory and phonetic perception. In: Carlson R, Granstrom B, editors. *The representation of speech in the peripheral auditory system.* Amsterdam: Elsevier; 1982. pp. 9–26.
2. Liberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition* 1985; 21:1–36.
3. Näätänen R, Lehtoskoskl A, Lennes M, Cheour M, Huotilainen M, Ilvonen A et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 1997; 385:432–434.
4. Kraus N, McGee T, Sharma A, Carrell TD, Nicol T. Mismatch negativity event-related potential elicited by speech stimuli. *Ear Hear* 1992; 13: 158–164.
5. Sittiprapaporn W, Tervaniemi M, Chindaduangratn C, Kotchabhakdi N. Preattentive discrimination of across-category and within-category change in consonant-vowel syllable. *NeuroReport* 2005; 16:1513–1518.
6. Sharma A, Dorman MF. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am* 1999; 106:1078–1083.
7. Näätänen R. Attention and brain function. 1992.
8. Winkler I, Lehtokoski A, Alku P, Vaiiio M, Czigler I, Csepe V, et al. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Brain Res Cogn Brain Res* 1999; 7: 357–369.
9. Okazaki S, Kanoh S, Takaura K, Tsukada M, Oka K. Change detection and difference detection of tone duration discrimination. *NeuroReport* 2006; 17:395–399.
10. Kujala T, Mllyviita K, Tervaniemi M, Alho K, Kallio J, Näätänen R. Basic auditory dysfunction in dyslexia as demonstrated by brain activity measurements. *Psychophysiology* 2000; 37:262–266.
11. Tervaniemi M, Castaneda A, Knoll M, Uther M. Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *NeuroReport* 2006; 17:1225–1228.
12. Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, et al. Auditory Cortex accesses phonological categories: an MEG mismatch study. *J Cognitive Neurosci* 2000; 12:1038–1035.
13. Sharma A, Kraus N, McGee T, Carrell T, Nicol T. Acoustic vs. phonetic representation of speech stimuli as reflected by the mismatch negativity event-related potential. *Electroencephalogr Clin Neurophysiol* 1993; 88: 64–71.
14. Maiste AC, Wiens AS, Hunt MJ, Scherg M, Picton TW. Event-related potentials and the categorical perception of speech sounds. *Ear Hear* 1995; 16:68–90.
15. Savela J, Kujala T, Tuomainen J, Ek M, Aaltonen O, Näätänen R. The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Brain Res Cogn Brain Res* 2003; 16:250–256.

16. Liberman AM, Harris KS, Hoffman H, Griffith B. The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 1957; **54**:358–368.

17. Massaro D. Categorical partition: a fuzzy-logical model of categorization behavior. In: Harnad S, editor. *Categorical perception: the groundwork of cognition*. Cambridge, UK: Cambridge University Press; 1997. pp. 254–283.

18. McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. *Cognition* 2002; **86**: B33–B42.

19. Andruski JE, Blumstein SE, Burton M. The effect of subphonetic differences on lexical access. *Cognition* 1994; **52**:163–187.

20. Klatt DH. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am* 1980; **67**:971–995.

21. Näätänen R, Gaillard AW, Mantysalo S. Brain potential correlates of voluntary and involuntary attention. *Prog Brain Res* 1980; **54**:343–348.

22. Alain C, Woods DL, Knight RT. A distributed cortical network for auditory sensory memory in humans. *Brain Res* 1998; **812**:23–37.

23. Jaramillo M, Ilvonen T, Kujala T, Alku P, Tervaniemi M, Alho K. Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Brain Res Cogn Brain Res* 2001; **12**:459–466.

24. Korpilahti P, Lang H, Aaltonen O. Is there a late-latency mismatch negativity (MMN) component? *Electroencephalogr Clin Neurophysiol* 1995; **95**:P96.