Phoneme discrimination and mismatch negativity in English and Japanese speakers

Marie D. Bomba¹, David Choly², and Elizabeth W. Pang^{1,2}

¹Program in Neurosciences and Mental Health, Sick Kids Research Institute, Toronto, Ontario, Canada

²Division of Neurology, Hospital for Sick Children, Toronto, Ontario, Canada

Abstract

Neural templates for phonemes in one's native language are formed early in life; these can be modified but are difficult to form *de novo*. These can be examined with mismatch negativity. Three phonemic contrasts were presented to adult native English compared to Japanese speakers who acquired English later: vowels native to both languages (/i/iy/); consonant-vowel contrasts (/da/wa/) phonemic in both languages; and consonant-vowel contrasts phonemic in English but not in Japanese (/ra/la/). For vowels, no MMN differences were found. For /da/wa/, MMN amplitude was significantly reduced in Japanese speakers. For /ra/la/, only 50% of the Japanese group showed an identifiable MMN. This suggests that phonemic templates are formed early in life and non-native consonant contrasts are difficult to learn later.

Keywords

mismatch negativity (MMN); event-related potentials; phoneme processing; second language learning

INTRODUCTION

While infants are born with the ability to perceive sounds from many different languages, adults have great difficulty discriminating speech not inherent to their mother tongue. In fact, it has been demonstrated that by 6 months of age, infants have difficulty discriminating non-native vowels [1], and by 12 months of age, they have difficulty discriminating non-native consonants [1,2]. The ability to distinguish subtleties and nuances within one's first language continues to be refined with experience, but the innate ability to distinguish foreign, unfamiliar language sounds diminishes, which renders second-language learning in adults a difficult task. The effects of language experience on speech perception are presumed to be due to neural coding of the acoustic components that are critical to native-language processing [2]. Neurophysiological studies using the mismatch negativity (MMN) have provided further evidence for language-specific "memory traces" in both infant [3,4] and adult listeners [5–8].

MMN amplitude is a measure of auditory perceptual disparity and an absent MMN indicates an inability to perceive any difference between sounds [10]. Furthermore, words with meaning, which are easily differentiable, elicit an enhanced MMN as compared to the MMN response elicited by pseudowords (e.g., [11]). With these well established characteristics, the MMN is a reliable neurophysiological marker of speech discrimination and language perception [9].

The MMN reflects memory traces specific to the mother tongue [5], and these are formed as a result of early exposure to one's native language environment [3]. Native speech contrasts elicit larger MMNs than sounds that do not belong to the native language [3,5,6]. For instance, in a study of Hungarian subjects, an MMN was not observable to vowel contrasts presented in Finnish, while Hungarians who were fluent in Finnish demonstrated an MMN to this contrast [6]. Native and non-native Estonian speakers demonstrated smaller amplitude MMNs in the non-native group for the non-native vowels than for the native contrasts, independent of the contribution of acoustic differences between vowels [5]. Similar results have been found in both adult [6] and infant [3] non-native speakers. Using magnetoencephalography (MEG), Menning [12] found larger and faster MMF responses to contrasts embedded in Japanese words in native Japanese speakers compared to native German speakers. However, this finding may have been confounded by the meaningfulness of the words in Japanese, which has been found also to foster the MMN response (e.g., [11]). Furthermore, both Japanese and American speakers demonstrated an MMN to contrasts that were phonemic in both languages (/ba/ /wa/), but the Americans showed a clearer MMN to the /ra/ /la/ contrast, which is phonemic in English but not in Japanese. This was interpreted as an indication that Japanese listeners were less sensitive to the /ra/ /la/ difference than American listeners [13]. This phenomenon of absent phonological differentiation has been well documented behaviourally [14].

The MMN also reflects the level of exposure within one's native language as well as changes in exposure that occur with the learning of a non-native language stimulus. The former was demonstrated in a study where native Japanese speakers showed larger MMNs to a less typical Japanese vowel (/e/ /ö/) than to a more common Japanese vowel (/e/) [15]. The latter has been demonstrated in a number of studies that showed differences in the MMN to newly learned non-native contrasts [12], and this was seen even in young Finnish children who developed an MMN to French vowels within two months of exposure in a French-speaking daycare centre [16].

Interestingly, most of the MMN studies have involved vowel stimuli and these have shown high concordance. However, studies involving consonant stimuli have not demonstrated the same consistency. Some studies have reported absent MMNs to non-native consonant contrasts [17], while others observed longer latencies [7], and still others have found no changes in amplitude [7,18]. From a neuropsychological perspective, the discordant results between vowels and consonants are not surprising as there is evidence indicating that consonants and vowels are functionally distinct and engage different neuronal systems [19].

Another explanation for the discordant consonant results may be that these studies fail to account for the fact that consonants fall along a spectrum, in terms of their phonemic

properties, and are thus processed differently. For instance, stop consonants (e.g., /ba/ /ta/) are processed categorically, which means that each stop consonant phoneme is perceived as a different category, resulting in more accurate identification and better discrimination [14]. Further along the spectrum, the stop-glide (e.g., /da/ /wa/), followed by the liquid (e.g., /ra/ /la/) consonants, are processed less categorically until we reach vowels at the other end of the spectrum which are not perceived categorically at all, but are processed in a continuous or acoustic mode [20]. In previous studies, the comparison of MMNs elicited by across- and within- category changes within the same language has revealed that the MMN is enhanced by across-category changes in vowels [6], place of articulation in consonants [17], and voice onset time (VOT) of stop consonants [8].

In this study, we wanted to investigate two different phenomena. First, we wished to support the MEG studies with electrophysiological evidence demonstrating an MMN to /ra/ /la/ in English, but not Japanese, speakers. Second, we wished to use the MMN to compare the processing of glide and liquid consonants, as well as vowels, in native and non-native English speakers. We specifically chose glide consonants and vowels that were phonemic in both the English and Japanese languages but chose liquid consonants (/ra/ /la/) that were phonemic in English but not in Japanese. Our use of stimuli along the phonemic spectrum allowed us to determine whether differences in the degree of categorical processing were reflected in the MMN in both native and non-native speakers.

METHODS

Sixteen healthy adults volunteered as participants and gave their informed consent. Subjects were either native English speakers (n=8, 4F, age range: 18–31 y, $\mu = 24$ y) or native Japanese speakers (n=8, 5F, age range: 23–37 y, $\mu = 30$ y) reported learning English after the age of 12 years.

Stimuli consisted of vowels and consonant-vowel syllables presented in separate sequences. For all sequences, 1000 stimuli were presented randomly with the “standard” stimulus occurring 85% of the time, and the “deviant” occurring with a probability of 15%. In the vowel condition, the standard was the English vowel /iy/ and the deviant was the English vowel /i/. The consonant-vowel syllables consisted of /da/ paired with /wa/, and /ra/ with /la/. /da/ and /ra/ were presented as the standards and /wa/ and /la/ were presented as deviants, in their respective conditions. All speech sounds were recordings of a female voice (250 ms duration) and presented at an ISI of 600 ms. All stimuli were presented at 72 dB SPL from a single loudspeaker placed at a distance of 1.1 m from the subject. Presentation order of conditions was randomized.

The EEG was recorded from 26 electrodes referenced to Cz and an averaged reference re-computed off-line. Data were recorded continuously using a Neuroscan 3.1 system with a sampling rate of 500 Hz, a bandpass of 1–30 Hz, and a gain of 1000x. Impedances were maintained below 5 k Ω .

Off-line, data were epoched into 500 ms intervals with a 100 ms pre-stimulus baseline. Data were filtered from 3–18 Hz. The responses evoked by the standard and deviant stimuli were

sorted and averaged by stimulus type. To find the MMN, the averaged responses evoked by the standard stimuli were subtracted from the responses evoked by the deviant stimuli. The MMN was picked as the largest negative peak between 150–300 ms. MMN amplitudes were measured from the averaged pre-stimulus baseline and latencies were measured from stimulus onset on a subset of fronto-central electrodes (F3, Fz, F4, C3, Cz, C4). For each condition, MMN amplitudes and latencies were subjected to 2-way repeated measures ANOVAs with native language and electrode as the factors and epsilon corrections applied.

RESULTS

The data for each of the stimulus conditions were analyzed separately using 2-way repeated-measures ANOVAs with native language (English vs. Japanese) and fronto-central electrodes (Fz, Cz, F3, F4, C3, C4) as the factors. For the vowel condition, there were no significant differences between native and non-native speakers on MMN latency and amplitude. For the /da/ /wa/ condition, MMN latency showed no significant differences between native and non-native speakers; however, MMN amplitude showed a main effect for language and a language by electrode interaction. Interestingly, for the /ra/ /la/ condition, the ANOVA could not be completed as an MMN was not observable for most of the non-native participants. The pattern of these findings are summarized in Table 1 and described in detail in the following sections.

/iy/ vs. /i/ condition

As expected, a clear MMN was seen, for the native English speakers, in the /i/ /iy/ condition, along the frontal chain (F3, Fz, F4) and vertex (Cz) electrodes at a latency of approximately 200 ms. As well, an MMN was observable for the Japanese speakers at these same electrodes. Figure 1a shows the grand-averaged response at Cz. The 2-way repeated measures ANOVA showed no significant differences between native English and native Japanese speakers for either MMN amplitude ($-0.62 \mu\text{V} \pm 0.08 \text{ SEM}$) or latency ($176.3 \text{ ms} \pm 3.6 \text{ SEM}$).

/da/ vs. /wa/ condition

As expected, clear MMN responses were seen for the native English speakers, in the /da/ /wa/ condition, along the frontal chain (F3, Fz, F4) and vertex (Cz) electrodes with a latency of just greater than 200 ms. For the non-native English speakers, a comparable MMN was not observable at all of these electrodes. Along the frontal chain, the MMN was smaller in amplitude, yet observable, for the non-native group, whereas at the Cz electrode, the MMN was difficult to identify. Figure 1b shows the grand-averaged response at Fz.

For MMN amplitude, a 2-way repeated measures ANOVA revealed a significant main effect for language ($F_{(1,14)} = 5.206$; $p < 0.05$) and a significant language x electrode interaction ($F_{(5,70)} = 5.433$; $\epsilon = 0.59$; $p_{(\text{adjusted})} = 0.003$). Post-hoc analyses demonstrated that the main effect for language was due to the MMN amplitude being significantly larger in the native English speakers ($-0.64 \mu\text{V} \pm 0.10 \text{ SEM}$) than the non-native speakers ($-0.09 \mu\text{V} \pm 0.06 \text{ SEM}$). For the language x electrode interaction, post-hoc analyses indicate that the MMN at Fz ($p < 0.02$) and F3 ($p < 0.02$) were significantly different while the MMN at Cz showed a

trend towards significance ($p=0.057$). MMN latency analyses were not significant ($180.2 \text{ ms} \pm 4.5 \text{ SEM}$).

/ra/ vs. /la/ condition

The results from the /ra/ /la/ condition were surprisingly complicated. The MMN, even for the group of native English speakers, was of low amplitude compared to the other stimulus conditions; however, an MMN was identifiable. In fact, a clear, classic MMN response was seen in 75% of the native English subjects with the remainder showing poorer, but identifiable MMN-like responses (see Fig. 1c for an example). For this group, mean MMN amplitude and latency was $-0.45 \mu\text{V} \pm 0.1 \text{ SEM}$ and $197.6 \text{ ms} \pm 5.6 \text{ SEM}$, respectively.

On the other hand, for the native Japanese group, the MMN was absent in 50% of the subjects. This made peak-picking impossible. In another 38% of the subjects, the MMN was poorly delineated, of extremely low amplitude, and possibly delayed; for this reason, peak-picking was not possible for this group and no further analyses were undertaken. A clear MMN was only seen in one subject (see Fig. 1d for an example).

DISCUSSION

The present study examined how different types of vowels and consonant-vowel phonemes are processed in the brain, and whether differences between native and non-native speakers exist. For this purpose, phonemes representative of the English spectrum of phonological processing were chosen - specifically, the vowels /i/ /iy/, which are processed continuously, the stop-glide consonants /da/ /wa/, which are processed categorically, and the liquids /ra/ /la/, which fall somewhere in between.

One of our main findings was that the native English speakers had larger amplitude MMNs than the non-native English speakers in both consonant-vowel syllable conditions, whereas no differences were observed in the vowel condition. This contrast between vowel and consonant change processing was expected. Since the vowels were phonemic in both English and Japanese speakers, the MMN response was expected to be large and clear, which was the case. Also, vowels are learned earlier in life [20], so the clear MMN contrast might reflect an innate neural template of both groups to vowel stimuli.

For the /da/ /wa/ condition, the MMN in native speakers was enhanced compared to MMNs of non-native speakers. This was not expected as the stop-glide consonants are essentially similar to stop consonants and are processed almost categorically. Thus, we expected both English and Japanese speakers to process stop-glide consonants in a manner similar to stop consonants. However, for the Japanese group, their MMNs were smaller and not as well-formed as those seen in the English group. One possible explanation might be the relative properties of the Japanese and English languages. The Japanese language only permits stop-glides, which results in one phonemic neural “template” for Japanese speakers, whereas English allows stop-glide and stop-liquid combinations, which results in four possible templates for English speakers. It has been suggested that Japanese speakers learning English do not form new templates but formulate a compromise based on their original template [21]. This would mean that, despite achieving fluency with English, Japanese

speakers do not possess the distinct templates of native English speakers, and this poorer categorical distinction is reflected in the MMN. These findings suggest that early language experience may have reshaped the state of the neural system at a pre-attentive level, altering perception in a way that assists native-language processing [9].

Our finding that MMNs to /ra/ /la/ were mostly absent in the Japanese group is also consistent with the literature which reports that liquid consonants are absent in the Japanese language. The poor-to-non-existent MMNs are consistent with the literature demonstrating that the MMN reflects memory traces specific to the mother tongue [5], these are formed as a result of early exposure to one's native language environment [3], and that poor second language perceivers have poorer MMNs [22]. The poorer quality, but still identifiable, MMN in the native English group probably demonstrates that categorical perception elicits the clearest response; however, the response deteriorates with less categorical perception (as is the case with liquid consonants).

Our /ra/ /la/ results are in contrast to magnetic MMN results using MEG. A number of studies have put forward explanations for the /ra/ /la/ effect in Japanese native speakers including a backward masking effect of foreign consonants by subsequent vowels [23,24]. This group also reported a shorter latency MMN in the Japanese group to their /ra/ la/ contrasts. This is, however, counter-intuitive as MMN latency is inversely correlated with the degree of perceived difference between the contrasts; that is, the larger the perceived contrast, the shorter the MMN. A simple explanation for the shortened MMN to the non-native stimuli might be an effect of the inter-stimulus interval (ISI) used in that study. Typical MMN studies employ a very brief inter-stimulus interval, usually around 500 ms. The MEG study cited above used an ISI almost twice as long with a mean ISI of 1 sec, and the issue was raised by this group that their earlier MMN might simply be due to a second N1m. This is a plausible explanation but remains in need of empirical testing.

Future studies should include a more comprehensive array of phonemes. For example, a limitation of this study was the absence of an exclusively categorical stop-stop consonant pair to complete the phonological spectrum. Including phonemes representative of the entire spectrum would allow us to pinpoint where changes in processing are occurring and elucidate the relationship between these changes and the degree of phonological processing. This information would impact our understanding of native language acquisition and early-versus later- learning of a second language.

CONCLUSIONS

The present data show that vowels and consonants are processed differently in the brain as measured by the MMN. Furthermore, neural differences exist in phonemic processing between native and non-native English speakers depending on the level of exposure to the phoneme early in life.

Acknowledgments

The authors acknowledge Felipe Allendes for assistance with data analyses and Matt J. MacDonald for editorial assistance. This research was partially supported by a Canadian Institutes of Health Research (MOP-89961) to the last author.

Partial support from Canadian Institute for Health Research (CIHR MOP-89961)

References

1. Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992; 255:606–608. [PubMed: 1736364]
2. Werker JF, Tees RC. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Beh Dev*. 1984; 7:49–63.
3. Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J. Development of language specific phoneme representations in the infant brain. *Nat Neurosci*. 1988; 1:351–353.
4. Rivera-Gaxiola M, Silva-Pereyra J, Kuhl PK. Brain potentials to native and non- native speech contrasts in 7- and 11-month-old American infants. *Dev Sci*. 2005; 8:162–172. [PubMed: 15720374]
5. Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*. 1997; 385:432–434. [PubMed: 9009189]
6. Winkler I, Kujala T, Tiitinen H, Sivonen P, Alku P, Lehtokoski A, et al. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*. 1999; 36:638–642. [PubMed: 10442032]
7. Rivera-Gaxiola M, Csibra G, Johnson MH, Karmiloff-Smith A. Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Beh Brain Res*. 2000; 112:1–11.
8. Sharma A, Dorman MF. Neurophysiologic correlates of cross-language phonetic perception. *J Acoust Soc Am*. 2000; 107:2697–2703. [PubMed: 10830391]
9. 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*. 2003; 16:250–256.
10. Pulvermüller F, Kujala T, Shtyrov Y, Simola J, Tiitinen H, Alku P, et al. Memory traces for words as revealed by the mismatch negativity. *Neuroimage*. 2001; 14(3):607–616. [PubMed: 11506534]
11. Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*. 2001; 38:1–21. [PubMed: 11321610]
12. Menning H, Imaizumi S, Zwitterlood P, Pantev C. Plasticity of the human auditory cortex induced by discrimination learning of nonnative, mora-timed contrasts of the Japanese language. *Learning and Memory*. 2002; 9:253–267. [PubMed: 12359835]
13. Zhang Y, Kuhl PK, Imada T, Kotani M, Tohkura Y. Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*. 2005; 26:703–720. [PubMed: 15955480]
14. Strange WH. Evolution of language. *J A M A*. 1984; 252 (8):1009.
15. Ikeda K, Hayashi A, Hashimoto S, Otomo K, Kanno A. Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neurosci Lett*. 2002; 321(3):133–136. [PubMed: 11880190]
16. Cheour M, Shestakova A, Alku P, Ceponiene R, Näätänen R. Mismatch negativity shows that 3–6-year-old children can learn to discriminate non-native speech sounds within two months. *Neurosci Lett*. 2002; 325 :187–190. [PubMed: 12044652]
17. Dehaene-Lambertz G, Dehaene S. Speed and cerebral correlates of syllable discrimination in infants. *Nature*. 1994; 370:292–295. [PubMed: 8035876]
18. Sams M, Aulanko R, Aaltonen O, Näätänen R. Event-related potentials to infrequent changes in synthesized phonetic stimuli. *J Cog Neurosci*. 1990; 2:344–357.

19. Caramazza A, Chialant D, Capasso R, Miceli G. Separable processing of consonants and vowels. *Nature*. 2000; 403:428–430. [PubMed: 10667794]
20. Fry DB, Abramson AS, Eimas PD, Liberman AM. The identification and discrimination of synthetic vowels. *Language and Speech*. 1962; 5:171–189.
21. Broselow E, Finer D. Parameter setting in second language phonology and syntax. *Second Language Research*. 1991; 7:35–59.
22. Diaz B, Baus C, Escera C, Costa A, Sebastián-Gallés N. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proc Natl Acad Sci*. 2008; 105(42):16083–16088. [PubMed: 18852470]
23. Koyama S, Gunji A, Yabe H, Yamada RA, Oiwa S, Kubo R, et al. The masking effect in foreign speech sounds perception revealed by neuromagnetic responses. *NeuroReport*. 2000; 11:3765–3769. [PubMed: 11117487]
24. Koyama S, Akahane-Yamada R, Gunji A, Kubo R, Roberts TP, Yabe H, et al. Cortical evidence of the perceptual backward masking effect on /l/ and /r/ sounds from a following vowel in Japanese speakers. *NeuroImage*. 2003; 18:962–974. [PubMed: 12725771]

Figure 1. Grand averaged MMN waveforms in the (a) /i/ /iy/ and (b) /da/ /wa/ conditions. Representative MMN for a ‘good MMN’ subject and a ‘poor MMN’ subject from the (c) English and (d) Japanese groups for the /ra/ /la/ conditions.

Table 1

Comparison of MMN results for English vs. Japanese speakers by condition

	MMN latency	MMN amplitude
/i/ vs. /iy/	n.s.	n.s.
/da/ vs. /wa/	n.s.	interaction: language x electrode
/ra/ vs. /la/	MMN absent in Japanese	MMN absent in Japanese

n.s. = not significant