Segmental and suprasegmental features in speech perception in Cantonese-speaking second graders: An ERP study

XIUHONG TONG, CATHERINE McBRIE, CHIA-YING LEE, JUAN ZHANG, LAN SHUAI, URS MAURER, AND KEVIN K. H. CHUNG

The Institute of Linguistics, Academia Sinica, Taipei, Taiwan
The Institute of Linguistics, Academia Sinica, Taipei, Taiwan
Faculty of Education, University of Macau, Macau, China
Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland, USA
Department of Psychology, University of Zurich, Zurich, Switzerland
Department of Special Education and Counselling, Hong Kong Institute of Education, Hong Kong, China

Abstract

Using a multiple-deviant oddball paradigm, this study examined second graders’ brain responses to Cantonese speech. We aimed to address the question of whether a change in a consonant or lexical tone could be automatically detected by children. We measured auditory mismatch responses to place of articulation and voice onset time (VOT), reflecting segmental perception, as well as Cantonese lexical tones including level tone and contour tone, reflecting suprasegmental perception. The data showed that robust mismatch negativities (MMNs) were elicited by all deviants in the time window of 300–500 ms in second graders. Moreover, relative to the standard stimuli, the VOT deviant elicited a robust positive mismatch response, and the level tone deviant elicited a significant MMN in the time window of 150–300 ms. The findings suggest that Hong Kong second graders were sensitive to neural discriminations of speech sounds both at the segmental and suprasegmental levels.

Descriptors: Speech perception, Cantonese, Segmental, Suprasegmental, MMN, p-MMR

The speech signal consists of segmental and suprasegmental information (Miller, 1978). In linguistics, the segmental features of speech are defined as “any discrete unit that can be identified, either physically or auditorily, in the stream of speech” (Crystal, 2003, pp. 408–409), such as consonants and vowels, which occur in a distinct temporal order. The suprasegmental features are concerned with units such as tone, stress, pitch, and length intonation, which always accompany the production of segmental features (Fox, 2000). Recently, there has been increasing neurophysiological research examining the neural discrimination of speech sounds at both segmental and suprasegmental levels in both adults and children (e.g., Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, & Gandour, 2007; Gomes et al., 1999; Korpilahti, Krause, Holopainen, & Lang, 2001; Maurer, Bucher, Brem, & Brandeis, 2003; Näätänen, Jacobsen, & Winkler, 2005; Näätänen et al., 1997; Näätänen, Paavilainen, Rinne, & Alho, 2001; Shafer, Morr, Kreuzer, & Kurtzberg, 2000; Shafer, Schwartz, & Kurtzberg, 2004). In adults, converging evidence shows that perception of both segmental and suprasegmental information is highly automatic, which can be reflected by the mismatch negativity (MMN), a component of the event-related potential (ERP). Specifically, the MMN is believed to be a marker of neural discrimination of speech perception at both the segmental level and the suprasegmental level (for a review, see Näätänen et al., 2007). However, the nature of the neurophysiological mechanisms of speech perception at both segmental and suprasegmental levels remains unclear in children thus far (Lee et al., 2012).

Notably, few studies have investigated the neural discrimination of different speech contrasts at both the segmental and suprasegmental levels in the same group of school-aged children. In particular, little is known about the neural correlates of different speech information in Chinese children, whose native language is a tonal language, which differs from English, a nontonal language (e.g., Lee et al., 2012; Meng et al., 2005). In the present study, we were interested in Hong Kong Chinese second graders’ brain responses to different speech contrasts at both the segmental and suprasegmental levels. Specifically, we recorded children’s brain activity in response to place of articulation and voice onset time (VOT) of an initial consonant in a consonant-vowel (CV) Chinese syllable, reflecting segmental perception, as well as Cantonese lexical tones including level tone and contour tone, reflecting suprasegmental perception. A multiple-deviant oddball paradigm was adopted in the present study. We aimed at investigating the question of whether Chinese second graders were sensitive to neural discrimination of changes in consonant and lexical tones of
a CV Chinese syllable in speech perception. That is, we tested whether the MMN, which is usually elicited in the oddball paradigm in adult ERP speech research, would be a marker of neural discrimination of speech perception at both the segmental and suprasegmental levels in Chinese second graders and whether it would show peak latencies and polarity directions as typically observed in adult speech perception.

**MMN in Adults**

The MMN is a negative-going and attention-independent component of the ERPs to any sufficiently discriminable change in a sequence of repetitive auditory stimuli, and it peaks between 150 to 200 ms with a distribution over the frontocentral sites but with inversion in polarity at inferior-posterior sites in adults (e.g., Chandrasekaran, Gandour, & Krishnan, 2007; Cheour, Leppänen, & Kraus, 2000; Näätänen, Pakarinen, Rinne, & Takegata, 2004; for a review, see Näätänen et al., 2007). Adult MMN studies show that MMN can be elicited by changes in acoustic features of an auditory stimulus that the participant can also discriminate behaviorally such as deviation in frequency (e.g., Jacobsen & Schröger, 2001; Paavilainen, Simola, Jaramillo, Näätänen, & Winkler, 2001), intensity (e.g., Kisley, Noecker, & Guinther, 2004; Näätänen et al., 2004), or duration (e.g., Grimm, Widmann, & Schröger, 2004; Jaramillo, Paavilainen, & Näätänen, 2000; Roebel, Widmann, & Schröger, 2003). The MMN can also be elicited by changes in more complex speech stimuli (e.g., changes in VOT, place of articulation, and lexical tone; e.g., Näätänen, 2001; Shafer et al., 2004; Sharma & Dorman, 1999; Tsang, Jia, Huang, & Chen, 2011). The MMN is believed to reflect automatic detection of auditory change at the preattentive stage and to originate from the auditory cortex (for a review, see Näätänen, Jacobsen, & Winkler, 2005). The increase of the MMN amplitude with a decrease of its peak latency reflects participants’ discrimination of auditory elements (for a review, see Näätänen et al., 2007).

The magnitude of the MMN is found to be sensitive to the degree of the difference between standard and deviants. Previous research on MMN has demonstrated that the peak amplitude of MMN is larger and it peaks earlier with an increase in the degree of deviance between standards and deviants (e.g., Dehaene-Lambertz, 1997; Pakarinen et al., 2009; Pakarinen, Takegata, Rinne, Huotilainen, & Näätänen, 2007; Tiitinen, May, Reimikainen, & Näätänen, 1994). The MMN is also modulated by the acoustic salience of a sound. That is, a large MMN tends to be elicited for a salient acoustic feature of a sound (e.g., Deouell, Bentin, & Giard, 1998; Isoguchi, Kanzaki, & Takahashi, 2011; Pellola et al., 2003); Deouell and colleagues (1998), for example, observed a larger MMN in peak amplitude for a change in frequency of a tone compared to the MMN found for changes in spatial location, intensity, and stimulus-onset asynchrony (SOA) even when other acoustic differences were matched across deviants. Pakarinen et al. (2007) reported a similar finding that frequency and duration elicited larger MMNs as compared to the intensity and location deviants. The authors also found that the MMNs elicited by duration peaked earlier and the MMNs elicited by intensity peaked later than MMNs elicited by frequency and location deviants, but the latencies of MMNs elicited by frequency and location did not differ from each other. The larger MMN suggests that it is much easier for speakers to detect the difference in frequency or duration between standard and deviant compared to the changes in intensity, spatial location, and SOA. Additionally, phonological status of a sound is another critical factor that affects the magnitude of the MMN, and this influence varies with listeners’ native language (for a review, see Dehaene-Lambertz & Gliga, 2004). Previous studies have shown that MMNs are easy to elicit with the same change within or across syllables in listeners for whom the change is phonemically relevant in their native language in comparison to listeners for whom the change is not phonemically relevant (e.g., Sharma & Dorman, 2000; Winkler et al., 1999).

**MMN in Children**

Attention is not necessary to elicit MMN, although some studies have demonstrated that MMN is somewhat modulated by attention (e.g., Hisagi, Shafer, Strange, & Sussman, 2010; for a review, see Sussman, 2007; Sussman, Winkler, & Wang, 2003). Thus, MMN may be a useful tool to examine auditory or speech perception in those who are limited in attention or motivation such as children (e.g., Friederici, Friedrich, & Weber, 2002; Kuhl, 1998; Lee et al., 2012; Shafer et al., 2000, 2010). There has been increasing research using the MMN paradigm to investigate auditory and speech perception in children. However, while MMN has been identified as a stable marker of neural discrimination for the different features of a sound in adults (for a review see Näätänen et al., 2007), the findings of change-related MMN have not been consistent across studies on auditory and speech perception in children thus far (e.g., Lee et al., 2012; Shafer, Yan, & Datta, 2010). Some researchers suggest that the polarity and latency of the MMN may vary with participants’ maturation (e.g., Lee et al., 2012; Maurer, Bucher, Brem, & Brandeis, 2003). For example, infants and young children tend to exhibit MMNs that have a longer duration or delayed onset compared to the MMN found in adults (e.g., Alho, Sainio, Sajaniemi, Reimikainen, & Näätänen, 1990: Cheng et al., 2013; Cheour et al., 1997; Cheour-Luhtanen et al., 1995; Lee et al., 2012). For example, Lovio et al. (2009) used a multideviant MMN paradigm to examine the neural response to changes in vowel, vowel duration, consonant, frequency (F0), and intensity in Finnish-speaking 6-year-old children. The authors found that significant adultlike MMNs were observed for all deviants against a standard, but the MMNs had a longer latency (mean 258–328 ms) compared to adults. However, some studies have shown no changes in MMN latencies from childhood to adulthood (e.g., Csepe, Dieckmann, Hoke, & Ross, 1992; Kraus et al., 1993).

A more dramatic age difference is the observation of a positive mismatch response (p-MMR) in infants and preschool children in response to discriminable changes in auditory and speech features of a sound. Infants and young children show a coexistence of the MMN and p-MMR in several studies (e.g., Cheng et al., 2013; Cheour et al., 1997; He, Hotson, & Trainor, 2009; Lee et al., 2012; Maurer et al., 2003; Shafer et al., 2010). Maurer and colleagues (2003), for example, used an oddball sequence with short intervals (every .38 s), comparing adults’ and 6- to 7-year-old children’s brain responses to small frequencies and phoneme deviance (standard: 1000 Hz and /ba/; larger deviance: 1060 Hz, /da/; smaller deviance: 1030 Hz, /da/; for frequency and phoneme, respectively). The authors observed a typical frontocentral MMN in the time window of 129–199 ms in adults, but a frontal p-MMR with anterior negativity in the time window of 179–207 ms was found in children. Their source localization analysis demonstrated that the p-MMR in the child group and MMN among the adults originated from superior temporal plane generators but with opposite polarity. They also observed that the p-MMR in children increased with the degree of phoneme deviance, but not with the degree of tone.
deviance. Thus, Maurer and colleagues suggested that the p-MMR found in children might have the same functional nature as the typical adult MMN, which may reflect additional or increased neural activation to deviants relative to standards.

In another recent study, Shafer et al. (2010) showed a coexistence of the MMN and p-MMR in 4- to 7-year-old children in response to an English vowel contrast, /ɪ/ versus /ɛ/. Specifically, the authors observed an adultlike MMN but in the time window of 300–400 ms in 4-, 5-, and 7-year-old children. In an earlier time window of 100–300 ms, they observed that children younger than 5.5 years old and some of the older children showed a significant p-MMR in response to the vowel contrast. The amplitude of the p-MMR was revealed to diminish in older children. The authors suggest that the p-MMR observed in younger children might reflect a recovery from refractoriness of the P100 (P1) obligatory response, and the disappearance of the p-MMR in older children might be attributed to “increased specificity of firing of neurons contributing to the P1 component, and to overlap with the N1b and MMN, as they mature” (p. 734).

To some extent, the findings of Maurer et al. and Shafer et al. were replicated in a more recent study of Chinese young children with Mandarin as a native language (Lee et al., 2012). Mandarin is a tonal language with a relatively simple syllable structure. A Mandarin syllable consists of onset, rime, and tone (McBride-Chang, Lin, Fong, & Shu, 2010). Using a multiple-deviant oddball paradigm, Lee and colleagues (2012) examined the developmental trajectory of lexical tone processing (standard: tone3 (T3), deviants: tone1 (T1) and tone2 (T2)), initial consonants (standard: /ba1/; deviants: /da1/, /ga1/), and vowels (standard: /dalu/; deviants: /di1/, /du1/) in three groups of Chinese children aged 4, 5, and 6 years. Lee et al. observed the coexistence of MMN and p-MMR in the same age group when responding to three types of speech features. More specifically, they found that the larger contrast T1/T3 elicited adultlike MMNs in the time window of 150–300 ms in the three age groups, but small contrast T2/T3 elicited p-MMRs in the 5- and 6-year-old groups. As for the vowel contrasts, the large contrast of /dalu/ with /da1/ elicited adultlike MMNs in the three age groups with a distribution in the bilateral frontocentral sites from 100–200 ms, whereas p-MMRs were observed for the small contrast of /di1/ with /da1/ in 5-year-old children with a distribution in the midline and right hemispheric sites. The p-MMR was larger over the left hemispheric recording sites from 150–200 ms; an adultlike MMN between 100–150 ms was also observed in 6-year-old children. As for the initial contrasts, p-MMRs were observed for the larger contrast /ga1/ and /ba1/ in the time window of 150–300 ms across three age groups. However, the small contrast /ba1/ versus /da1/ elicited p-MMR only in the 4-year-old group. These findings demonstrated that, with increasing age, the adultlike MMN becomes more prominent, whereas the p-MMR decreases, suggesting that the MMN may reflect enhanced and more mature discrimination ability, and the p-MMR may reflect the more difficult discrimination in young children. The authors thus suggest that MMN and p-MMR reflect different functional characteristics.

However, questions such as what the p-MMR reflects and when it may be present or absent continue to be debated among researchers (e.g., Cheng et al., 2013; Lee et al., 2012; Maurer et al., 2003). For example, some researchers claim that the p-MMR may act as an analogy of sorts to the adultlike P3a, which is suggested to reflect distractability or an involuntary attention shift or the automatic categorization of stimuli (e.g., Escera, Alho, Winkler, & Näätänen, 1998; He et al., 2009; Kushnirenko, Ceponiene, Baland, Fellman, & Näätänen, 2002; Shestakova, Huotilainen, & Cheour, 2003). Other researchers hold the view that the p-MMR might reflect a recovery from refractoriness, indexing the detection and encoding of the acoustic properties of a stimulus in connections in the primary auditory cortex (e.g., Dehaene-Lambertz & Dehaene, 1994; Friederici et al., 2002; Shafer et al., 2010; Shafer, Yu, & Garrido-Nag, 2012). Additionally, the p-MMR was observed in most prior studies in infants or preschool children. It is rarely observed in school children (e.g., Lee et al., 2012). Nevertheless, more research is needed to understand the functional significance of the mismatch response (positive or negative) in speech perception in children. In particular, neurophysiology studies of speech perception in Cantonese children are very limited at both the segmental and suprasegmental levels.

The Present Study

There were two purposes of the present study. The first purpose was to examine whether the MMN could be a neural index of Cantonese initial consonant and lexical tone discrimination in Cantonese-speaking second graders. The second purpose was to investigate whether the peak latency and polarity of the MMN would vary across different types of deviants. We manipulated the VOT and place of articulation of an initial consonant of Chinese single-syllable and lexical tones including level tone and contour tone. The initial consonant and lexical tone are two components of a Cantonese syllable. The consonant in a Cantonese syllable is optional. That is, syllables may consist of vowels only. In total, there are 20 initial consonants. As a tone language, Cantonese uses variations in pitch (e.g., high vs. low) or the pitch pattern (e.g., rising vs. falling) to distinguish lexical meanings of morphemes. Lexical tone is, therefore, obligatory for a Cantonese syllable (Tan, Ching, Chan, Cheng, & Mak, 1995). There are six distinctive tones (up to nine, depending on how one counts it): tone1–high level (55), tone2–high rising (25), tone3–midlevel (33), tone4–low falling (21), tone5–low rising (23), tone6–low level (22). Among the six tones, tone1, tone3, and tone6 are categorized as level tones in which there are few or no changes of pitch over time; tone2, tone4, and tone5 are contour tones—these pitches change appreciably over time (Wang, Yang, & Cheng, 2009).

In fact, there have been a number of studies using an ERP or magnetoencephalography (MEG) paradigm examining the neural correlates of place of articulation and VOT perception in adults in nontonal languages (e.g., Dehaene-Lambertz, 1997; Phillips et al., 1995, 2000; Shafer et al., 2004; Sharma & Dorman, 1999). In general, significant MMNs are elicited by changes of VOT both within and across consonants, as well as by changes in place of articulation in a typical MMN time window of 150–250 ms, and MMNs tend to be larger for the changes across a phonetic boundary than within a phonetic boundary (Sharma & Dorman, 1999). A few studies in lexical tone perception have also reported detectable MMNs for both level tone and contour tone in tonal languages (e.g., Cheng et al., 2013; Lee et al., 2012; Sittiprapaporn, Chindaduangratn, Tervaniemi, & Khotchabkhadi, 2003; Tsang et al., 2011; Xi, Zhang, Shu, Zhang, & Li, 2010).

However, to our knowledge, there are only two developmental studies so far examining the trajectory of the MMN in

---

1. Chao (1980) first transcribed lexical tone in a numerical notational system by using five levels (from lowest 1 to highest 5) to describe relative height, shape, and duration of pitch contour. High, middle, and low are the descriptions of pitch ranges. Level, rising, and falling are the specifications of the slopes of the pitch contours.
Chinese-speaking children. The first study was conducted by Meng et al. (2005). Using the MMN paradigm, Meng et al. (2005) examined the differences in brain responses to pure frequency changes, pure tone changes, an initial consonant change, and a lexical tone change in typically developing and dyslexic children aged 11 years. They found that all contrasts elicited significant MMNs in typically developing children, and the dyslexic group showed smaller MMNs in processing initial consonant and vowel syllable contrasts as well as marginally different MMNs in response to tone perception. The other study was that of Lee et al. (2012) as mentioned above. However, the native language for the participants in both the Meng et al. (2005) and Lee et al. (2012) studies was Mandarin. Although both Mandarin and Cantonese are tonal languages, they differ substantially in tone attributes. Specifically, there are only four tones in Mandarin, but there are up to nine tones in Cantonese.

The present study aimed to investigate Cantonese-speaking children’s preattentive brain responses to changes in consonants and lexical tones to determine the nature of Cantonese-speaking children’s neural discrimination of speech sound at both the segmental and suprasegmental levels.

Based on findings of prior research on consonant and lexical tone perception, we expected that changes in consonants and lexical tones would result in a significant either negative or positive mismatch response in Cantonese second graders. More specifically, we expected that all four of the deviants, that is, VOT, place of articulation, level tone, and contour tone, would elicit significant adultlike MMNs, and the p-MMN, which has been suggested to disappear after 7 years old, would not appear in second grade Cantonese-speaking children. In addition, we hypothesized that the brain responses to speech sounds for Cantonese second graders might peak later or less than those that emerged in previous studies of Mandarin-speaking children when considering the differences in tone attributes between Cantonese and Mandarin. For example, there are only four tones in Mandarin but up to nine tones in Cantonese, which may lead to greater density in perceptual space for Cantonese relative to Mandarin. There are also certain differences in acoustic features between Cantonese tones and Mandarin tones. For example, for Cantonese there are several acoustic correlates of lexical tones, that is, pitch height, pitch contour, and pitch onset. Recent evidence shows that pitch contour is important to distinguish tone contrasts differing in pitch contour such as high level (55) versus high-rising (25), whereas pitch onset (F0 onset) is more important to distinguish tone contrasts having the same contour but different pitch height, such as midlevel (33)-low level (22) (Tong, McBride, & Burnham, 2014). By contrast, pitch contour is the primary acoustic cue to distinguish Mandarin tone contrasts (Howie, 1976). In other words, Cantonese children’s perceptual system for Cantonese tones is much more fine grained and delicate relative to the perceptual system for Mandarin tones (see So & Best, 2010). Thus, the sophisticated tone system to which Cantonese-speaking children are exposed tunes them in to subtle acoustic details, such as F0 onset. For this reason, we would expect some perceptual differences between Cantonese- and Mandarin-speaking children.

Method

Participants

Participants were 17 Hong Kong second graders (7 females and 10 males) aged between 88 and 99 months (Mage = 93.59 months, SD = 3.52 months). All children were Cantonese-native speakers.

None of the participants had a history of neurological, psychiatric, brain injury, or hearing problems reported by parents.

Materials and Design

The multiple-oddball paradigm was adopted in the present study. The experimental stimuli were five Cantonese syllables: /gaa1/, /gaa2/, /gaa3/, /daa1/, and /kaa1/. The syllable /gaa1/ was used as the standard; the other four syllables were assigned as deviants. The stimuli were recorded by a female Cantonese speaker. All stimuli were kept at 75 dB SPL; F0 of the five syllables was between 200.4 and 288.3 Hz. The duration of the stimuli was normalized to 170 ms. Thus, in each trial, the stimuli lasted 170 ms, with a 600-ms interstimulus interval (ISI). Among the five stimuli, the contrasts /gaa1/ versus /daa1/, and /gaa1/ versus /kaa1/ were classified to reflect the difference in initial consonant defined as segmental level condition; the contrasts /gaa1/ versus /gaa2/, and /gaa1/ and /gaa3/ were categorized as reflecting the difference in lexical tones defined as the suprasegmental level condition. The contrast of /gaa1/ and /daa1/ represented the difference in place of articulation; that is, the second formants of the stimuli showed different start frequencies (the frequency of F1, F2, and F3 was 1336.5 Hz, 2033.3 Hz, and 3063.3 Hz, respectively, for /gaa1/ stimuli, and the frequency of F1, F2, and F3 was 1333.8 Hz, 2026.4 Hz, and 3049.5 Hz, respectively, for /daa1/ stimuli), and the pitch was matched between /gaa1/ and /daa1/. The contrast of /gaa1/ and /kaa1/ differed in the VOT with a 41.3 ms difference. The contrast of /gaa1/ and /gaa2/ differed in pitch contour, and the /gaa1/ and /gaa3/ contrast differed in pitch height. The average pitch value of /gaa1/, /gaa2/, and /gaa3/ was 282.9 Hz, 236.0 Hz, and 221.7, respectively. In addition, the minimum and maximum pitch values for /gaa2/ and /gaa3/ were 200.4 Hz and 288.3 Hz, respectively. The acoustic feature of each contrast is shown in Figure 1.

Procedure

All participants were tested individually in a sound-attenuated ERP lab. Before the experiment, we presented caregivers who brought the child to our lab with a consent form and parental questionnaire and asked them to complete each before testing began. The parental questionnaire was used to screen out children with attention deficit hyperactivity disorder (ADHD), and other related learning disabilities; this questionnaire also ensured that the first language of all children was Cantonese. Following completion of these forms, the preparation of the electroencephalographic (EEG) recording began. After preparation, one Cantonese-speaking helper, who had been trained before the testing, instructed each participant on details of the experimental procedure.

Participants were seated in a comfortable chair in front of a computer monitor at a distance of 80 cm from the computer monitor. The participants were required to watch a movie, “The Mole,” in silence while listening to the experimental stimuli. The stimuli were binaurally delivered via headphones. The stimuli were presented through three blocks, each of which consisted of 515 trials, starting with 15 trials of standard, followed by 40% deviants (10% for each deviant), and 60% standard. Thus, there were 150 trials for each type of deviant in total. Among all stimuli including all deviant trials and standard trials, 80% of the stimuli had tone1 but with a different onset, that is, either with /g/ or /k/ or /d/; 80% of these had the /g/ onset. The order of the trials within each block...
was pseudorandomized. Participants were given 2 min for a break between blocks during the experiment.

**EEG Recordings**

EEG activity was recorded from a 64-channel (Ag-AgCl) Neuroscan system. The 64 channels were arranged according to the 10/20 system with the reference electrode located between Cz and CPz. The electrode located between Fpz and Fz served as the ground electrode. The vertical electrooculogram (EOG) was obtained from below versus above the left eye, and the left versus right lateral orbital rim defined the horizontal EOG. During recording, the electrode impedances were kept below 15 kΩ. The EEG and EOG signals were amplified with a band-pass of 0.05–70 Hz and digitized at a sampling rate of 1000 Hz.

**Data Analysis**

We analyzed the EEG data offline using Scan 4.4 software. The EEG data were referenced to the average of all electrodes (Lehmann & Skrandies, 1980). Given that there are no inactive locations on the head, choosing the average reference avoids biasing the analysis by the selection of a particular electrode as reference (Koenig & Gianotti, 2009). The continuous data were filtered with a 0.10–30 Hz band-pass. The filtered data were segmented into epochs of 700 ms long, with a 100-ms prestimulus interval used for baseline correction. The first 15 standard trials and epochs with artifacts exceeding ±100 μV were discarded (e.g., Cheng et al., 2013; Lee et al., 2012). Also, the standard trials that immediately followed the deviant trials were excluded from averaging. The mean numbers of accepted deviants were 122, 123, 124, and 124 for deviants of place of articulation, VOT, level tone, and contour tone, respectively, and that of the accepted standards was 238.

The mean amplitudes of each condition were computed in the epoch from 150–300 ms and 300–500 ms at the electrodes F3, Fz, F4, FC3, FCz, and FC4. Repeated measures analyses of variance (ANOVAs) with experimental condition (standard, deviants), site (frontal, centrofrontal), and hemisphere (left, middle, right) as within-subject factors were performed separately for the segmental level condition and the suprasegmental level condition in the two time windows. To examine whether each deviant elicited significant MMNs, the planned comparisons were performed between each deviant condition and the shared standard condition. For each ANOVA, the Greenhouse-Geisser adjustment to the degrees of freedom was applied.
freedom was used to correct for the violations of sphericity associated with repeated measures.

Results

The mismatch responses at Fz and the topographic voltage maps obtained by subtracting the standard from the four types of deviants and difference mismatch potential maps are shown in Figures 2 and 3 for all types of stimuli, respectively. As shown in Figure 2, there was no obvious MMN or p-MMR for place of articulation or contour tone deviants in the time window from 150–300 ms. However, the level tone deviant elicited a clear MMN in the time window of 150–300 ms, whereas a p-MMR was observed in the VOT deviant. In addition, there appear to be clear MMNs in the time window from 300–500 ms for each of the deviants.

Segmental-Level Conditions (/daa1/ and /kaa1/ versus /gaa1/ standard)

150–300 ms time window. Statistical analyses revealed that there was a significant main experimental condition effect in this time window, $F(2,32) = 14.38, p < .001$, $\eta^2_p = .47$. Planned comparisons demonstrated that the VOT deviant (/kaa1/) was more positive than the standard ($M = 1.00 \mu V, SD = .29$, and $-1.12 \mu V, SD = .30$, respectively, $p < .001$). However, there was no significant difference found between the place of articulation deviant (/daa1/) and

Figure 2. ERP waveforms to the standard and deviant stimuli for (a) placement of articulation, (b) VOT, (c) level tone, and (d) contour tone at the electrode of Fz.

Figure 3. Maps display the topographic distribution of the mean amplitude in the MMN analysis window for the deviant minus standard difference for (a) placement of articulation, (b) VOT, (c) level tone, and (d) contour tone.
the standard \((M = -36 \mu V, SD = 0.28, -12 \mu V, SD = 0.30,\) respectively, \(p = .18\)).

**300–500 ms time window.** A significant main effect of experimental condition was found in this time window, \(F(2,32) = 9.48, p < .01, \eta^2_p = .37\). Follow-up pairwise comparisons indicated that the mean amplitude of the place of articulation deviant was more negative than the mean amplitude of the standard \((M = -2.85 \mu V, SD = 0.34, -2.02 \mu V, SD = 0.20,\) respectively, \(p < .01)\), and the mean amplitude of the VOT deviant was more negative than the mean amplitude of the standard \((M = -3.48 \mu V, SD = 0.33, -2.02 \mu V, SD = 0.20,\) respectively, \(p < .001)\).

**Suprasegmental-Level Conditions (/gaa2/ and /gaa3/ versus /gaa1/ standard)**

**150–300 ms time window.** For the lexical tone contrasts, a significant main effect of experimental condition was found in this time window, \(F(2,32) = 8.66, p < .01, \eta^2_p = .35\). Follow-up pairwise comparisons indicated that the mean amplitude of the level tone deviant was more negative than the mean amplitude of the standard \((M = -2.86 \mu V, SD = 0.27, -1.12 \mu V, SD = 0.30,\) respectively, \(p < .05)\), but there was no significant difference found between the contour tone deviant and standard \((M = 1.51 \mu V, SD = 0.27, -1.12 \mu V, SD = 0.30,\) respectively, \(p = .25)\). In addition, a significant interaction effect of Site \(\times\) Condition was found in this time window, \(F(2,32) = 3.36, p < .05, \eta^2_p = .17\). Follow-up analysis showed that the mean amplitude of the level tone deviant was more negative than the standard in the frontal site \((p < .01)\), and there was only a marginally significant MMN found for the level tone in the frontal central site \((p = .06)\).

**300–500 ms time window.** In this time window, the ANOVA performed on the mean amplitude of the MMNs revealed that there was a significant main effect of experimental condition, \(F(2,32) = 4.80, p < .05, \eta^2_p = .23\). Planned comparisons showed that the level tone deviant was more negative than the standard \((M = -2.89 \mu V, SD = 0.37, -2.02 \mu V, SD = 0.20,\) respectively, \(p < .01)\). The mean amplitude elicited by the contour tone deviant was also more negative than the mean amplitude elicited by the standard \((M = -2.58 \mu V, SD = 0.31, -2.02 \mu V, SD = 0.20,\) for contour tone deviant and standard, respectively, \(p < .05)\). No hemisphere asymmetry effect was found in either of the two time windows across deviants for either the segmental and suprasegmental analyses in the present study.

**Discussion**

In this study, we used the ERP approach with an auditory multiple-deviant oddball paradigm to investigate Hong Kong second graders’ Cantonese speech perception at both the segmental and suprasegmental levels. Specifically, we recorded second graders’ brain responses to place of articulation and VOT at the segmental level, and to level tone and contour tone at the suprasegmental level. Interestingly, we found that the segmental level feature of VOT elicited a robust positive response (p-MMR) relative to the standard between 150 and 300 ms; in contrast, the suprasegmental feature of level tone elicited a significant negative response (typical MMN) at the same time window. No detectable p-MMR or MMN was found in the other two contrasts, that is, the place of articulation versus standard and the contour tone versus standard in the time window of 150–300 ms. However, in a later time window of 300–500 ms, all deviants elicited significant MMNs, presumably related to a late MMN (Korpilahti, Lang, & Aaltonen, 1995; Maurer et al., 2003) or the late discriminative negativity (Cheour, Korpilahti, Martynova, & Lang, 2001) as reported in previous studies. Our findings suggest that Cantonese-speaking second graders are sensitive to automatic neural discrimination of the acoustic differences between the standard and deviants including VOT, place of articulation, and lexical tones. We may also infer from our results that second graders might be more sensitive to the acoustic change in VOT compared to place of articulation, reflected by a robust p-MMR found in the VOT deviant in a relatively early time window.

Our finding that robust MMNs were elicited for Cantonese-speaking second graders in response to all deviants including the segmental level (i.e., VOT and place of articulation) and the suprasegmental level (i.e., level tone and contour tone in the time window of 300–500 ms) is consistent with results found in previous MMN research on speech perception in children (e.g., Maurer et al., 2003; Shafer et al., 2010). For example, Shafer et al. (2010) found that 4- to 5- and 6- to 7-year-olds showed a robust MMN in the time window between 300–400 ms in response to an English vowel contrast (e.g., /i/ versus /e/); they also found a p-MMR in all younger children and half of the older children in the time window of 100–200 ms, and the younger children showed a bit longer latency (up to 260 ms). However, the MMNs to all deviants found in our present study were a bit different from the MMN found in adults in terms of peak latency. That is, the MMNs found in our second graders peaked in a relatively late time window, namely, from 300–500 ms, whereas MMN is often found to peak in the time window between 150–250 ms in adults (for a review, see Näätänen et al., 2007). Previous research on MMN in adults suggests that the latency and amplitude of the MMN reflect the degree of participants’ discriminability of speech sounds. With the increasing difficulty between the standard and the deviant, “the MMN shifts later in time and becomes smaller in amplitude” (Shafer et al., 2010; p. 736). Thus, on the one hand, our finding suggests that Cantonese-speaking second graders show neural discrimination of speech sounds at both the segmental and suprasegmental levels; on the other hand, the MMN peaking in a relatively late time window may indicate that Cantonese-speaking second graders may still not be able to discriminate speech sounds automatically at the preattentive stage as accurately as adults. This seems consistent with findings from behavioral research that children performed worse than adults in discrimination of fine-grained differences in speech stimuli (Elliott & Hamner, 1988).

In accordance with the findings from Mandarin-speaking children of the ages of 4, 5 and 6 years by Lee et al. (2012), the VOT deviant elicited a detectable p-MMR in the time window of 150–300 ms in the present study. That is, p-MMR and MMN were coexistent in Cantonese-speaking second graders when discriminating the VOT deviant. Notably, the p-MMR is generally observed in oddball studies of newborns, infants, and young children when responding to auditory or speech stimuli (e.g., Alho et al., 1990; Cheng et al., 2013; Korpilahti et al., 2001; Lee et al., 2012; Maurer et al., 2003). For example, Kurtzberg, Vaughan, Kreuzer, & Flieglter (1995) reported a p-MMR in newborns when responding to a pitch change among stimuli presented with 750 and 1,000 ISIs. Morr, Shafer, Kreuzer, & Kurtzberg (2002) observed larger p-MMRs to two types of pitch deviants in infants between 2 and 12 months using a 750 ms ISI. In a very recent study, Cheng et al. (2013) found a p-MMR in newborns, but an adultlike MMN was observed in infants from the age of 6 months.
when responding to a larger contrast between Mandarin lexical tone1 and tone3, whereas the small contrast between lexical tone2 and tone3 was only elicited in infants but not in newborns. Some studies have shown a robust p-MMR in young children (e.g., Lee et al., 2012; Maurer et al., 2003), and a few studies have shown a p-MMR in children aged 7 years in relation to auditory and speech perception (e.g., Shafer et al., 2000). Moreover, a significant p-MMR was elicited by the vowel /i/ in Chinese tone1 and tone4 in Mandarin-speaking participants aged between 18 to 24 years in a recent study (Kuo et al., 2013). Thus, the existence of the p-MMR is not a phenomenon that is specific to infants and preschoolers only.

However, there is still a debate regarding the nature of the p-MMR in the literature on auditory and speech perception (Lee et al., 2012; Maurer et al., 2003; Shafer et al., 2010). One explanation for the presence of the p-MMR is that the p-MMR may reflect the immature neural change-detection processes in infants and young children (e.g., Lee et al., 2012; Maurer et al., 2003). Perhaps, with maturation, the p-MMR is progressively replaced by the MMN by about 4 years old. However, this explanation is somewhat questionable because the p-MMR has been observed in several studies in older children as well as in adults, as mentioned above (e.g., Lee et al., 2012; Maurer et al., 2003; Shafer et al., 2010; Kuo, Lee, Chen, Liu, & Cheng, 2013). Thus, the presence of the p-MMR found in the present study of Cantonese-speaking second graders with a mean age of 7.8 years when discriminating the VOT deviant may not be accounted for by the immature response hypothesis.

There are also some other different accounts proposed by researchers to explain the presence of the p-MMR. For example, the p-MMR might be an immature type of P3a. P3a is a component with a frontocentral scalp distribution and is suggested to reflect the process of involuntary attention shifting; it is usually present after the MMN component in research focused on MMN (e.g., He et al., 2009; Kushenerenko et al., 2002; Lee et al., 2012; Maurer et al., 2003). As argued by Shafer et al. (2010), if the processes indexed by P3a are dependent on discrimination carried out in the system indexed by MMN, the p-MMR, which is hypothesized to be an immature type of P3a, should be present after the MMN component. However, the p-MMR has been observed in an earlier time window preceding the MMN in several recent studies (e.g., Lee et al., 2012; Maurer et al., 2003; Shafer et al., 2010). The MMN also came first in the present study. Because of the debate regarding whether the elicitation of P3a requires MMN, the p-MMR found in the present study may not have the same function as the P3a. Another account, which we favor, is that the p-MMR may reflect the recovery from refractoriness of the P100 (P1) response with a source in auditory cortex in infants and young children (e.g., Mikkola et al., 2007; Shafer et al., 2010). This component might be canceled by the MMN for a specific contrast or at a specific time point (e.g., Lee et al., 2012; Shafer et al., 2010). We thus propose that the p-MMR found in the present study may be a recovery of P1. As P1 may reflect very early auditory sensory encoding, the p-MMR may index detection and encoding of the acoustic properties of a stimulus in the present study.

Another interesting finding in the present study is that the MMNs emerged continuously in the 150–300 ms and 300–500 ms time windows for the level tone (tone3), but they were only observed in the time window of 300–500 ms for the contour tone (tone2) in Cantonese-speaking second graders. This finding may suggest that the temporal course of the level tone and contour tone perceptions may be different from each other in Cantonese-speaking second graders. This may be because of the different attributes of level tone and contour tone in Cantonese. As shown in Figure 1, the pitch height was held constant throughout the duration for level tone3, but the pitch height was changed with the unfolding of time for contour tone2. In other words, the change of tone3 with time was static but the change of tone2 was dynamic. Previous behavioral studies on speech perception at either the segmental or suprasegmental levels have revealed that dynamic and static cues have different decay rates in auditory working memory, and a number of studies have shown that the perceptual trace of dynamic cues decays faster than that of static cues (e.g., Pisoni, 1973; Tong, Francis, & Gandour, 2008). The MMN is believed to reflect a comparison process between a memory constructed to the standard stimulus and to each incoming stimulus; thus, the MMN observed in the time window of 150–300 ms in the present study may reflect the fact that Cantonese-speaking second graders were able to detect the difference between tone3 and tone1 within a short period after the onset of the deviant, and they may monitor the process until the offset of the stimuli, which may be reflected by the MMN found in the time window of 300–500 ms. In addition, tone3 is a level tone, whereas tone2 is a rising tone in Cantonese. Behavioral research on tone acquisition and perception has suggested that the rising tone is more difficult to perceive than are the level tone and falling tone for Chinese (e.g., Li & Thompson, 1977; Snow, 1998; Varley & So, 1995). Thus, the absence of a significant MMN for tone2 in a relatively early time window (i.e., 150–300 ms) also suggests that children may have difficulty in processing tone2, which is a rising tone, leading to no detectable MMN found in a relatively early time window.

It is worth noting that, to some extent, Cantonese-speaking children appear to show brain responses that are similar to those of Mandarin-speaking children, at least as demonstrated by Lee et al. (2012), to speech sounds at both the segmental and suprasegmental levels. For example, as in the present study, Lee et al. (2012) found that Mandarin-speaking children at 4, 5, and 6 years old were able to discriminate changes in initial consonants, vowels, and lexical tones, reflected by either positive or negative mismatch responses. However, some differences were also revealed in brain responses to changes in initial consonants and lexical tones between Mandarin-speaking and Cantonese-speaking children. For example, the coexistence of the p-MMR and MMN was found for Cantonese-speaking children for the VOT deviant, but only the p-MMR was found for the Mandarin-speaking children. In addition, only MMNs were found in tone perception for Cantonese-speaking children. In contrast, Lee et al. (2012) found a coexistence of the p-MMR and MMN in Mandarin young children in tone perception. Also, the MMNs found in Cantonese-speaking second graders appear to peak a bit later compared to the MMNs found in Mandarin-speaking young children for tone perception. For example, MMNs were found in the window of 150–200 ms in 6-year-old children in the study by Lee et al. for the contrast between tone1 and tone2.

Two possibilities may account for those differences. First, as discussed above, the presence and/or absence of the p-MMR might be associated with a change of age. Thus, age differences for the participants between the two studies may be one of the factors leading to those differences. In our study, children were those with a mean age of 7.8 years old. In contrast, the participants were children 4, 5, and 6 years old in Lee et al.’s study (2012). A second reason may be the result of the differences in
the phonological systems between Cantonese and Mandarin. For example, our finding that MMNs for our Cantonese-speaking participants peaked a bit later compared to the MMNs found in young Mandarin-speaking children for tone perception may be accounted for by the differences between Mandarin and Cantonese tones. Cantonese has more tones than Mandarin. Thus, the perceptual space may be less dense in Mandarin relative to Cantonese tones. In addition, Cantonese tones also differ from Mandarin tones in acoustic features. For example, in Cantonese, tone3 is the midlevel tone in which the pitch height is held constant, but the tone3 in Mandarin is a low dipping tone in which the pitch height changes with the unfolding of time. We propose that the differences in perceptual space and acoustic features between Cantonese and Mandarin might also result in differences in neural discrimination in speech information between Cantonese and Mandarin speakers.

In fact, developmental studies have demonstrated that Cantonese children acquire all tones later than Mandarin-speaking children do. Thus, Cantonese-speaking second graders with a mean age of 7.8 years old may still not be able to discriminate speech sounds completely automatically at the preattentive stage. This hypothesis is consistent with the evidence from behavioral studies on the development of Cantonese tone perception showing that Cantonese children achieve adultlike performance in lexical perception by the age of 10 (Ciocca & Lui, 2003).

Conclusions

Our study was among the first to provide neuropsychological evidence for speech perception at both the segmental and suprasegmental levels in Cantonese-speaking children. Results confirmed that Cantonese-speaking second graders were sensitive to neural discriminations of changes in VOT, place of articulation, and lexical tones, including both level and contour tones, which were reflected by the robust adultlike MMNs, but with a relatively longer latency found in all deviants. In addition, the p-MMR was observed in those children who were more than 7 years old. This may suggest that the perception of the phonological contrasts used in our present study is not mature by 7 years of age for Cantonese speakers. This component may reflect a recovery from refractoriness of the P1 component (Shafer et al., 2010). However, there is still a debate regarding the nature of the p-MMR in the literature on auditory and speech perception (Lee et al., 2012; Maurer et al., 2003). Future work might attempt to replicate and extend our findings on Cantonese speech perception in children of different ages. Future research should also help in further determining when and what factors may affect the presence or attenuation/absence of the p-MMR in children. Moreover, our finding that the significant MMN could be found in both early and late time windows for level tone but only in the late time window for contour tone may suggest that Hong Kong Cantonese-speaking second graders show differential sensitivities to pitch height and pitch contour in tone perception.

References


