Sensory Processing of Linguistic Pitch as Reflected by the Mismatch Negativity

Bharath Chandrasekaran, Ananthanarayan Krishnan, and Jackson T. Gandour

Objective: To assess the extent to which acoustic and phonetic change-detection processes contribute to the mismatch negativity (MMN) to linguistic pitch contours.

Design: MMN was elicited from Mandarin and English speakers using a passive oddball paradigm. Two oddball conditions were constructed. In one condition (T1/T2i), the Mandarin high-level tone (T1) was compared with a convex high-rising tone (inverted T2, henceforth referred to as T2i) that occurs as a contextual variant of T1 in running speech. In the other (T2/T2i), the concave high-rising tone (T2) was compared with T2i. Phonetically, T1/T2i represents a within-category contrast for native speakers, whereas T2/T2i represents a between-category contrast. The between-category pair (T2/T2i), however, is more similar acoustically than the within-category pair (T1/T2i). In an attention-demanding behavioral paradigm, the same speakers also performed an auditory discrimination task to determine the perceptual distinctiveness of the two tonal pairs.

Results: Results revealed that the Chinese group, relative to the English, showed larger MMN responses and earlier peak latencies for both conditions, indicating experience-dependent enhancement in representing linguistically relevant pitch contours. At attentive stages of processing, however, the Chinese group was less accurate than the English in discriminating the within-category contrast (T1–T2i).

Conclusions: These findings demonstrate that experience-dependent neural effects at early preattentive stages of processing may be driven primarily by acoustic features of pitch contours that occur in natural speech. At attentive stages of processing, perception is strongly influenced by tonal categories and their relations to one another. The MMN is a useful index for examining long-term plasticity to linguistically relevant acoustic features.

(Ear & Hearing 2009;30;552–558)

INTRODUCTION

The traditional mismatch negativity (MMN) is an automatic change-related frontocentrally distributed evoked cortical potential with a passive oddball paradigm. In a passive oddball paradigm, participants are actively involved in a distracter task, typically reading a book or watching a movie, while they listen to repetitive auditory stimuli. To obtain the auditory MMN, the brain's response to a frequently presented "standard" is subtracted from the response to a rarely presented "deviant." The resulting "difference waveform" reflects the extent to which the brain detects a change in the incoming auditory stimulus stream (Näätänen 2001; Näätänen et al. 2007). The MMN has been used extensively in basic and clinical research as an objective index of central sound representation and auditory discrimination (Kraus et al. 1995; Kraus & Cheour 2000; Kujala & Näätänen 2001; Bishop 2007; Kujala et al. 2007).

The MMN has also been shown to be influenced by long-term experience and short-term training. Cross-language MMN experiments have demonstrated experience-dependent modulation of segmental information (consonants, vowels) (Sharma & Dorman 2000; Näätänen 2001) and suprasegmental units (vowel quantity, lexical tones) (Nenonen et al. 2005; Tervaniemi et al. 2006; Chandrasekaran et al. 2007c) in speech. Short-term training paradigms have demonstrated an increased magnitude of the MMN as a function of training (Kraus et al. 1995; Tremblay et al. 1997; Kujala et al. 2001; Cheour et al. 2002). Based on the studies of segmental information in speech, it has been proposed that in addition to a language-universal acoustic change detection process, MMN responses of native speakers are modulated by a left-hemisphere–dominant language-specific phonetic change-detection process (Näätänen 2001; Näätänen et al. 2007). Because untrained non-native speakers do not have long-term stored representations for native phonemes, the model predicts that they rely exclusively on the acoustic change-detection process. The processing advantage for native speakers, therefore, is hypothesized to result from the additional influence of stored representations via the phonetic change-detection process (Näätänen et al. 2007). Similarly, processing advantages posttraining have also been hypothesized to result from the phonetic change-detection process (Tremblay et al. 1997).

However, it has been shown that even as early as the auditory brain stem, native speakers exhibit more accurate representation of linguistically relevant pitch contours (Krishnan et al. 2005, 2009). At the level of the cerebral cortex, the preattentive MMN response to segmental (Joanisse et al. 2007) and suprasegmental (Chandrasekaran et al. 2009) stimuli may be strongly influenced by sensory factors. Moreover, early preattentive processing of linguistic pitch has been demonstrated to be lateralized to the right hemisphere in native speakers (Luo et al. 2006), a finding that conflicts with the idea that MMN responses to native categories are lateralized to the left hemisphere (Näätänen et al. 2007). Luo et al. proposed a two-stage model in processing linguistic pitch contours: a preattentive stage of processing involved in pitch analysis in the right hemisphere and an attentive stage of processing driven by higher-level linguistic representations in the left hemisphere. Indeed, we have suggested that language experience may even shape basic acoustic processes that underlie the acoustic change-detection process (Tervaniemi et al. 2006; Chandrasekaran et al. 2007b,c). Perhaps native speakers also benefit from superior acoustic representations of the standard and deviant traces, which result in a more robust change-detection response relative to non-native speakers.

The aim of this cross-language (Chinese, English) study was to examine the extent to which acoustic and categorical processes contribute toward processing of nonspeech homologues of Mandarin Chinese tonal contours at preattentive and attentive stages of processing. MMN and behavioral accuracy measures were obtained from responses to three nonspeech

Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, Indiana.

0196/0202/09/3005-0552/0 • Ear & Hearing • Copyright © 2009 by Lippincott Williams & Wilkins • Printed in the U.S.A.

stimuli. Two passive oddball sequences were designed to distinguish acoustic from phonetic processing. In one sequence (T2/T2i), the concave high-rising tone (T2) was compared with an inverted mirror image of itself (T2i). In the other (T1/T2i), the high-level tone (T1) was compared with T2i, a convex high-rising tone that occurs in connected speech as a contextual variant of T1 (Xu 1997, 2006). The pair T1 and T2i represents a within-category contrast; T2 versus T2i, a between-category contrast. In terms of acoustics, however, T2 is more similar to T2i, the between-category pair, than T1 is to T2i, the within-category pair.

By virtue of language experience, we expected the MMN to be larger for Chinese than English participants, regardless of categorical status. For native speakers, if the MMN is driven exclusively by long-term stored categorical representations, we predicted that the between-category pair (T2/T2i) would show a larger MMN than the within-category pair (T1/T2i). The opposite result was predicted if it were driven predominantly by acoustics (T1/T2i > T2/T2i). Furthermore, by using a control group of English speakers who do not have long-term stored representations of Mandarin tones, we can determine the extent to which processing at early cortical stages is influenced by the existence of categories. With respect to behavioral discrimination, we predict that native speakers of Mandarin, relative to English speakers, will be less accurate in discriminating the within-category pair (T1/T2i), suggesting a categorical bias.

In contrast, we predict superior accuracy for native speakers in discriminating the across-category pair (T2/T2i). By comparing preattentive and attentive processing of linguistic pitch contours, we are able to determine the extent to which categorical status influences the two stages of processing.

SUBJECTS AND METHODS

Subjects

Eleven native speakers of Mandarin Chinese (5 men, 6 women) and 11 adult native speakers of American English (5 men, 6 women) participated in the evoked response potential (ERP) experiment. The two groups were closely matched with respect to age (Chinese: mean = 24.4 yr, SD = 4.2 yr; English: mean = 26.2 yr, SD = 3.2 yr) and number of years of formal education (Chinese: mean = 17.4 yr, SD = 2.3 yr; English: mean = 18.4 yr, SD = 2.4 yr). Participants were strongly right-handed (> 90%) as measured by the Edinburgh handedness inventory (Oldfield 1971). All participants were nonmusicians as determined by a musical history questionnaire (Wong & Perrachione 2007). None of the native Mandarin or American English speakers had more than 3 yr of formal training in music or any combination of musical instruments, and none had any type of musical training within the past 5 years. In addition, participants completed a language history questionnaire (Li et al. 2006). Native speakers of Mandarin Chinese had no English instruction before the age of 11 yr, and none of the American English speakers had any previous exposure to Mandarin or any other tone language. Both groups had normal-hearing sensitivity (i.e., pure-tone air conduction thresholds of 25 dB HL or better in both ears) at frequencies of 0.5, 1.2, and 4 kHz. Participants were paid for their participation and gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

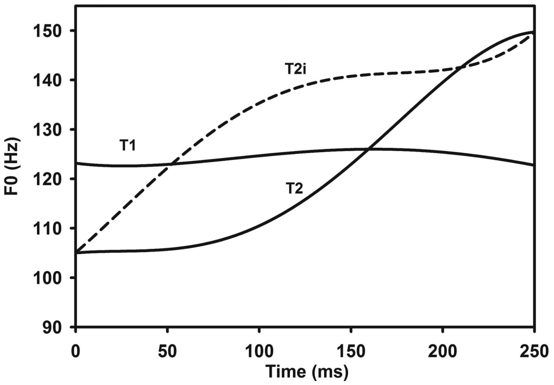

Fig. 1. Voice fundamental frequency (f_0) contours of the three IRN stimuli. T1 and T2 (solid lines) are modeled after average f_0 contours of time-normalized Mandarin lexical tones that occur on monosyllabic words produced in citation form (Swaminathan et al. 2008). T2i (dashed line) represents an f_0 contour that does not exist in citation form but does occur as a contextual variant of T1 in connected speech. T2i was the deviant in both ERP conditions. T1 was standard in one condition (T1/T2i) and T2 in the other (T2/T2i). The f_0 onset, offset, and direction of movement (rising) are identical for T2 and T2i. They differ primarily on the basis of overall contour (T2, concave; T2i, convex). T1 and T2, on the other hand, differ in terms of onset, offset, direction, and contour. T1 and T2 represent the high-level and -rising lexical tones of Mandarin; T2i represents the inverse of T2, a convex, curvilinear f_0 contour that does not occur on tones produced in isolation. ERP, evoked response potential.

Stimuli

Three time-varying nonspeech (iterative ripple noise [IRN]) f_0 contours were created using procedures described by Swaminathan et al. (2008). To generate IRN stimuli, white noise is added to itself with a delay. It has been shown that with repeated iterations, a percept of pitch emerges that is consistent with a pitch of the inverse of the delay used. The pitch percept increases in saliency with the number of delay and add procedures (Yost 1996a,b). In the current study, a high iteration step (32) was used for all contrasts with gain set to 1. At 32 iterations, IRN stimuli exhibit clear bands of energy at the fundamental and its harmonics, but unlike speech, they are devoid of formant structure. Of the three time-varying f_0 contours (Fig. 1), two (T1 and T2) were modeled after naturally occurring citation forms of Mandarin lexical tones using a fourth-order polynomial equation (Xu 1997). T1 and T2 reflect native Mandarin high-level and -rising tones, respectively, differing from each other on the basis of onset, offset, height, direction, and shape of the f_0 contour. The third stimulus (T2i) was created by flipping the T2 contour around a best-line fit of the original T2 polynomial. Thus, T2i has the same onset, offset, and direction as T2. Pitch trajectories consistent with T2i do not occur in isolated monosyllables but can occur as a result of tonal coarticulation in connected speech as an allophonic variant of the high-level tone (T1) (Xu 1997, 2006). The duration of all four stimuli was fixed at 250 msec, including the 10-msec rise and fall ramps to reduce spectral splatter. Waveforms and narrowband spectrograms of the stimuli are displayed in Figure 2.

ERP Experiment

Data acquisition • Participants sat in an acoustically and electrically shielded booth facing an LCD monitor. They were instructed to ignore the sounds presented binaurally via insert earphones and watch a self-selected movie with subtitles. Partic-

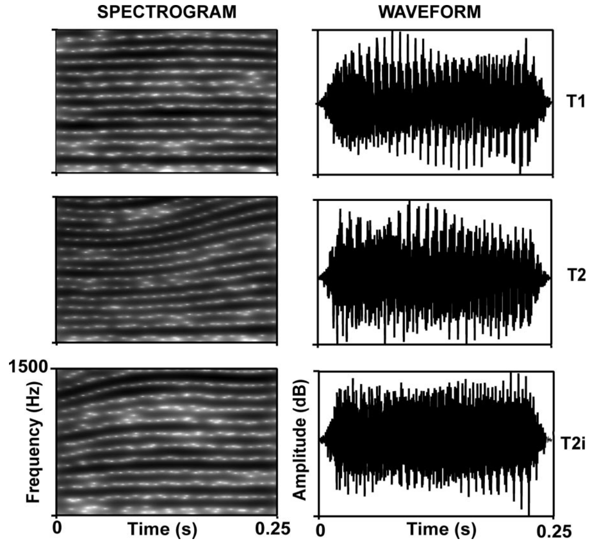

Fig. 2. Narrowband spectrograms and waveforms of the three nonspeech (IRN) stimuli used in this study. The three stimuli were created at a high-iteration step (32) from broadband noise. In the spectrograms, clear bands of energy are evident at the time-varying f_0 (first dark band) and its harmonics, but unlike speech, IRN stimuli show no formant structure. As evidenced in the waveform representations, the three stimuli have relatively flat amplitude contours, allowing amplitude to be controlled independently of pitch contour. IRN, iterated rippled noise.

ipants were also instructed that they would have to provide a synopsis of the movie at the end of the experimental session. The same set of instructions was provided to both groups in English. For all the oddball sequences, the interstimulus interval (onset to onset) was fixed at 667 msec. Within an oddball sequence, the frequent stimulus (standard) occurred at a probability of 0.85 and the infrequent stimulus at a probability of 0.15. The order of presentation of stimuli was pseudorandom; i.e., at least one standard stimulus preceded the deviant.

The ERP experiment consisted of four oddball sequences. In one condition (T1/T2i), T2i (convex curvilinear rising) was presented as the deviant and T1 (high level) as the standard. In a second condition, T2 (concave curvilinear rising) was presented as the standard, with T2i as the deviant. Thus, the two sequences had a common deviant (T2i) occurring in the context of either T1 or T2. In the two remaining sequences, the oddball sequences were reversed with T2i as the common standard with T1 and T2 separately occurring as the deviants. Data collection continued until 100 artifact-free deviants were collected for each sequence. The experiment ran for approximately 2 hr including subject preparation. All stimuli were generated and controlled by a signal generation and data acquisition system (Intelligent Hearing Systems, Smart EP). Stimuli were presented binaurally at 75 dB SPL through magnetically shielded insert earphones (Bio-logic, ER3A, TIP 300 tubal insert earphones).

Evoked-potential recording • For each subject, silver chloride electrodes were mounted on frontal midline (Fz) and central midline (Cz) and the linked mastoid electrode sites. These electrode locations were chosen because the MMN is known to be maximal at frontal areas and shows a reduction in amplitude at more central locations. The MMN, unlike the N2b, is known to invert at the mastoids. By examining polarity reversal at the linked mastoids, we were able to confirm whether the negativity at Fz and Cz truly reflected a mismatch response. The tip of the nose served as the reference electrode and the forehead as the ground. The responses were recorded at a 1000 Hz sampling rate. Electrode impedance was kept below 5000 ohms. Electrodes monitoring vertical eye movements were used in removing eye blink artifacts, as defined by epochs with voltage changes exceeding 60 μV. The electrical signal was band-pass filtered at 1 to 30 Hz.

Data analysis • The baseline for the grand averaged waveforms was defined as the average of the amplitude values between −100 and 0 msec (onset of stimuli). To obtain the MMN, the deviant waveforms from the T1/T2 and T2/T2i sequences were subtracted from standard waveforms presented in the T2i/T1 and T2i/T1 sequences, respectively. Subtracting the deviant from the same stimuli presented as the standard effectively controls for any acoustical differences between stimuli. The MMN peak latency was calculated as the most negative voltage in the MMN window between 125 and 300 msec. The MMN mean amplitude was calculated as the mean voltage from a 100-msec window centered on the MMN peak latency. The mean amplitude measure is a more robust index relative to peak amplitude measures (Luck 2005; Bishop 2007). In previous studies, this index has consistently reflected experience-dependent effects related to the MMN (Chandrasekaran et al. 2007b,c, 2009).

The MMN mean amplitude and peak latencies were analyzed using a three-way mixed model analysis of variance (ANOVA) for the effects of group (Chinese, English), condition (T1/T2i, T2/T2i), and location (Fz, Cz). In the model, subjects (random) were nested within group, a between-subject factor; condition and location (fixed) were within-subject factors.

Behavioral Experiment

Experimental protocol • All participants were asked to perform speeded-response discrimination judgments of the three IRN stimuli (T1, T2, and T2i) immediately after the ERP experiment. They were first presented with a practice set of stimuli to gain familiarity with the task. All pairs of stimuli (same and different trials) occurred once in the practice set. Each trial consisted of a pair of stimuli separated by a 300-msec interstimulus interval (onset to onset). The “same” and “different” trials had equal probability of occurrence. All trials were randomized within each block. The two stimuli within different trials were also presented in a random order. Subjects were asked to press the left (same) or right (different) mouse button to indicate their discrimination judgment during the 1.5 sec response interval after each pair. Stimuli were presented binaurally by means of computer playback (E-Prime) though a pair of Sony MDR-7506 headphones at a comfortable listening level (72 dB SPL).

Data analysis • Response accuracy (%) was calculated for each subject, and tonal pair was used in the ERP experiment (T2–T2i, T2i–T2, T1–T2i, T2i–T1). The different pairs were pooled across order of presentation of tones within pairs, yielding 20 trials per pair (T2–T2i, T1–T2i). Arcsine-transformed proportions of correct responses (Winer et al. 1991) were subjected to a mixed-model ANOVA to examine whether discrimination accuracy

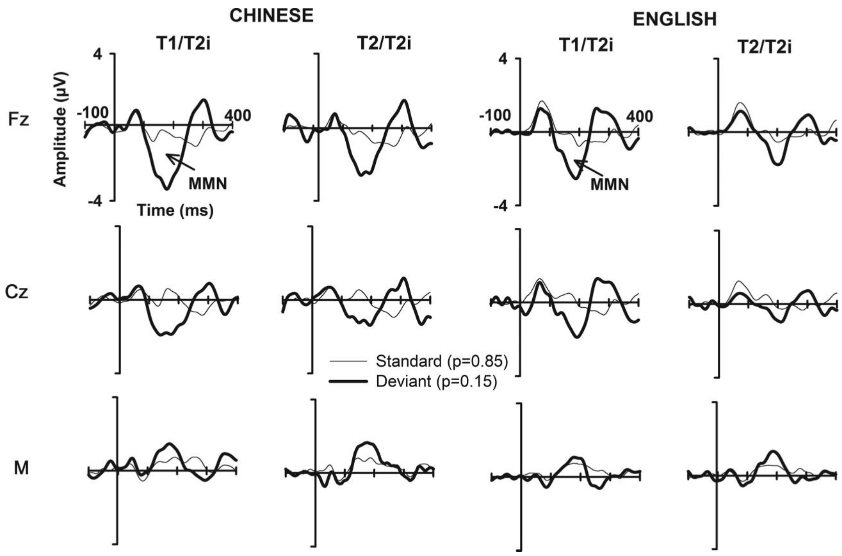

Fig. 3. Grand average standard ($p = 0.85$) and deviant ($p = 0.15$) waveforms per group (Chinese, English) and condition (T1/T2i, T2/T2i) at three electrode locations (Fz, Cz, mastoid). Regardless of group or condition, the MMN was larger at Fz than at Cz and showed the typical polarity reversal at the mastoids. With respect to group, MMN-related negativity was larger for the Chinese relative to the English group across both conditions. For both groups, the T1/T2i condition elicited more robust MMN responses relative to the T2/T2i condition. The MMN peaked earlier for the Chinese group relative to the English group regardless of condition.

varied as a function of group (between subjects: Chinese, English) and tonal pair (within subjects: T2–T2i, T1–T2i).

RESULTS

ERP Experiment

The grand average waveforms for the two groups (Chinese, English), two conditions (T1/T2i, T2/T2i), and three locations (Fz, Cz, linked mastoids) are shown in Figure 3. Both conditions (T1/T2i, T2/T2i) elicited robust MMN responses for both groups within the 125 to 300 msec time window. The MMN was reduced in amplitude at Cz, relative to Fz, and reversed in polarity at the mastoid location, indicative of a "true" mismatch response.

Results from the omnibus three-way ANOVA (group × condition × location) for mean amplitude revealed significant main effects of group ($F_{1,20} = 22.07$, $p = 0.0001$, $\eta^2_{\text{partial}} = 0.52$), condition ($F_{1,20} = 20.63$, $p = 0.0002$, $\eta^2_{\text{partial}} = 0.51$), and location ($F_{1,40} = 56.57$, $p < 0.0001$, $\eta^2_{\text{partial}} = 0.59$). None of the two- or three-way interaction effects between group, condition, and location reached statistical significance.

The MMN mean amplitude for each group (Chinese, English) and condition (T1/T2i, T2/T2i) at the electrode location Fz is displayed in the left panel of Figure 4. Pooling across conditions and locations, post hoc Tukey-adjusted comparisons revealed that the Chinese group had a larger MMN mean amplitude than the English group ($t_{20} = 4.70$, $p = 0.0001$). Pooling across groups and location, post hoc Tukey-adjusted comparisons revealed that the T1/T2i condition elicited larger MMN mean amplitude responses relative to the T2/T2i condition ($t_{20} = 4.54$, $p = 0.0002$). Pooling across groups and conditions, post hoc Tukey-Kramer–adjusted comparisons revealed that the MMN mean amplitude at Fz was significantly greater than Cz ($t_{40} = 3.51$, $p = 0.0008$).

The mean peak latency for each group and condition at the electrode location Fz is plotted in the right panel of Figure 4. Results from the omnibus ANOVA yielded a significant main effect of group ($F_{1,20} = 10.89$, $p = 0.003$, $\eta^2_{\text{partial}} = 0.35$). The main effect of condition or location failed to reach significance. Although a trend was evident, the interaction effect between group and condition failed to reach significance ($F_{1,20} = 3.47$, $p = 0.08$). None of the two- or three-way interactions reached significance. Pooling across condition and location, MMN peak latency for the Chinese group was significantly earlier than the English group ($t_{20} = -20.56$, $p = 0.0036$).

Behavioral Experiment

A two-way (group × tonal pair) repeated-measures ANOVA conducted on the arcsine-transformed proportion of

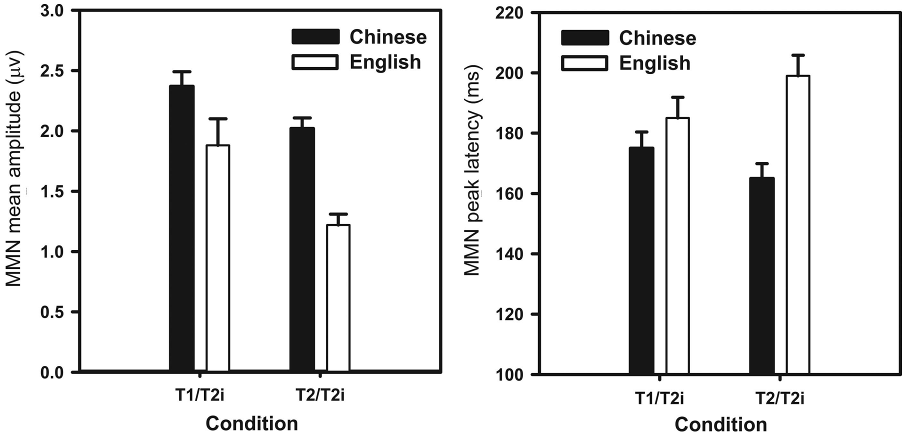

Fig. 4. Mean MMN amplitude (left panel) and peak latency (right panel) values for the two groups (Chinese, English) per condition (T1/T2i, T2/T2i) as measured from the Fz electrode location. Regardless of condition, the mean MMN amplitude was greater for the Chinese group relative to the English group. Per condition, the mean MMN amplitude was greater for the T1/T2i condition relative to the T2/T2i condition. Regardless of condition, the MMN peaked earlier for the Chinese group relative to the English group.

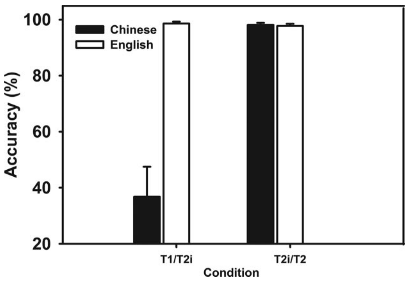

Fig. 5. Discrimination accuracy as a function of paired f_0 contours (T1–T2i, T2–T2i) and group (Chinese, English). Both groups performed at ceiling level for T2–T2i. For T1–T2i, Chinese participants were less accurate than the English. Only the Chinese group was significantly less accurate for T1–T2i relative to T2–T2i.

correct responses yielded significant main effects of group ($F_{1,20} = 48.07$, $p < 0.0001$, $\eta^2_{\text{partial}} = 0.71$) and tonal pair ($F_{1,20} = 65.15$, $p < 0.0001$, $\eta^2_{\text{partial}} = 0.77$) and a significant interaction effect between group and tonal pair ($F_{1,20} = 52.23$, $p < 0.0001$, $\eta^2_{\text{partial}} = 0.723$). The percent correct per group (Chinese, English) and tonal pair (T1–T2i, T2–T2i) are displayed in Figure 5. Per group, only the Chinese exhibited significant differences in response accuracy between pairs ($F_{1,20} = 117.02$, $p < 0.0001$, $\eta^2_{\text{partial}} = 0.85$; English: $F_{1,20} = 0.36$, $p = 0.56$). Per tonal pair, only T1–T2i yielded a group difference in response accuracy. The Chinese group was less accurate than the English in T1–T2i ($F_{1,20} = 99.81$, $p < 0.0001$, $\eta^2_{\text{partial}} = 0.83$). T2–T2i was not significantly different between groups ($F_{1,20} = 0.08$, $p = 0.78$).

DISCUSSION

The major findings from the current cross-language study suggest that language experience modulates the automatic early cortical processing of native pitch contours. Modulation of the MMN, an index of early cortical processing, seems to be sensitive to acoustic dimensions, whereas behavioral discrimination in native speakers seems to be exclusively determined by categorical effects. Chinese speakers, compared with English speakers, exhibit larger MMN responses and earlier peak latencies across conditions (T1/T2i, T2/T2i). However, both groups show a larger MMN in response to the T1/T2i than the T2/T2i condition. The presumed reason is that T2/T2i varies on fewer pitch dimensions or has smaller physical difference relative to T1/T2i. Specifically, T2 and T2i have similar onsets, offsets, and pitch direction (rising). In contrast, T2i and T1 differ on the basis of a number of pitch dimensions, including onsets, offsets, direction, overall slope, and pitch height. For the Chinese group, MMN responses are less robust in the between-category (T2/T2i) than the within-category condition (T1/T2i). This finding suggests that early preattentive cortical processing of linguistic pitch contours is strongly sensitive to acoustic properties relative to tonal categories for native speakers. But at attentive stages of processing, categories strongly influence discrimination of native pitch contours. The Chinese group is less accurate than the English in discriminating between the within-category contrasts (T1–T2i). Only the Chinese group is less accurate in discriminating the T1–T2i contrast relative to T2–T2i, which suggests a reduced perceptual saliency for the allophonic contrast (T1–T2i) at attentive stages of processing. Because both groups perform at ceiling levels in discriminating the T2–T2i contrast, it is difficult to determine whether native speakers have an advantage in processing the between-category pair.

Acoustic Versus Categorical Basis for MMN Modulation to Linguistic Pitch Contours

Our findings suggest that the MMN may serve as an index of acoustic features that are differentially weighted by language experience (Chandrasekaran et al. 2007a,b,c). In the current study, experience-dependent effects for a pitch contour (T2i) that occurs in natural speech as an allophonic variant of T1 are elicited even when T2i is compared with T1. Such findings support an acoustic basis for experience-dependent effects at early cortical stages of processing linguistically relevant pitch contours. Sensory analysis at the early cortical stages of processing may be integrated into a semantic representation at later attentive stages of processing. Indeed, lexical tones have been shown to elicit a stronger preattentive MMN response from native speakers of Mandarin in the right hemisphere than in the left (Luo et al. 2006), whereas consonants evoke the opposite asymmetry (Shtyrov et al. 2000; Luo et al. 2006). Lexical tones and consonants are distinct acoustically. The former are characterized by slowly changing spectral variation and the latter by rapidly changing temporal variation. Such hemispheric asymmetries suggest that early cortical processing of linguistic pitch is driven primarily by acoustic features before being mapped onto a semantic representation at later stages of processing.

Our data fit this two-stage model insofar as we see a strong acoustic basis for MMN modulation and a strong categorical basis for discrimination accuracy. Chinese speakers show a more robust MMN response than English speakers for the T1/T2i condition. Yet, they are less accurate than English speakers in distinguishing perceptually between these two pitch contours. This disparity between electrophysiology and behavior suggests that later attention-modulated stages of processing are influenced by long-term representations of tonal categories. Similarly, it has been suggested that the MMN evoked by consonants can be strongly influenced by sensory factors beyond what is predicted by overt categorization and discrimination judgments (Joanisse et al. 2007). Similarly, in the musical domain, relative to abstract musical categories, the MMN response has been shown to be highly sensitive to pitch dimension (Trainor et al. 2002; Fujioka et al. 2004, 2005). Taken together, these data suggest a strong acoustic bias for the MMN regardless of the nature of the stimulus, i.e., speech, nonspeech, or music. Thus, the large acoustic contribution to the MMN response may explain a recent study that has demonstrated enhanced MMN responses in musicians listening to non-native linguistic pitch contrasts (Chandrasekaran et al. 2009).

As far as the MMN itself is concerned, we are unable to completely disentangle sensory and phonetic contributions to the overall change-detection process. In a recent magnetoencephalographic investigation of automatic pitch-change detection, the MMNm (magnetic counterpart of the MMN) is seen to reflect both a purely sensory process and a more cognitive memory-based comparator process separated in time but overlapping in cortical space (Maess et al. 2007). The early part of the MMN (105 to 125 msec) is caused mainly by a noncomparison, sensory process. The later part, on the other hand, reflects a cognitive-based comparison between the extracted standard and the incoming deviant. In the current study, we restricted the analysis to a 125 to 300 msec time window to rule out contributions from the purely sensorial comparator mechanism (Maess et al. 2007). Despite this, we still find that the MMN is determined primarily by acoustic (sensory) features. Although we cannot completely rule out categorically based contributions in the current experiment, we speculate that the acoustic change-detection mechanism prevails in view of the disparity between MMN modulation and discrimination accuracy observed in the native group.

It is thus plausible that the acoustic change-detection process is shaped by language experience. Long-term experience with pitch contours can shape the neural response to pitch as early as the level of the auditory brain stem. For example, Mandarin speakers show more robust pitch representation than English speakers across the four Mandarin tones (Krishnan et al. 2005). Pitch encoding in response to Mandarin tones is more robust in monolingual, English-speaking musicians compared with nonmusicians (Wong et al. 2007). By using IRN stimuli, the MMN responses are not confounded by lexical-semantic interference, which could give native speakers an advantage (Pulvermuller et al. 2008). Moreover, our findings are in agreement with previous data on differential sensitivity to pitch dimensions extracted from IRN homologues of ecologically representative pitch contours (Chandrasekaran et al. 2007b).

Our results are of direct relevance in understanding the nature of experience-dependent effects at early cortical stages of processing. Recent studies have demonstrated that long-term auditory experience and short-term training can enhance the acoustic representation of linguistically relevant stimuli at the level of the brain stem (Krishnan et al. 2005; Song et al. 2008). In the current study, we show that the acoustic change-detection process is dominant in determining the degree of MMN modulation to linguistic pitch contours. It is therefore possible that experience-related changes to the MMN result from more efficient and enhanced acoustic representations at subcortical and cortical stages of processing.

Neurobiological Basis for Early Tuning of Auditory Information

In animals, it is well established that experience-dependent neural plasticity is not limited to the cerebral cortex (Suga et al. 2000, 2002, 2003). The cortex improves the signal quality of its input by shaping the brain stem response to any repeated stimuli via corticofugal feedback. This feedback mechanism is augmented for behaviorally relevant stimuli, thereby further enhancing the cortical response. A functional connectivity is also possible between the human brain stem and cortex. Abnormal brain stem timing in learning disabilities is related to a reduction in cortical sensitivity to acoustic change and deficient literacy skills (Banai et al. 2005). In tone languages, rapid f_0 movements are required for high intelligibility of contour tones (Abramson 1978). Using an unsupervised learning paradigm, it has been shown that the first derivative of f_0 (velocity) yields more accurate classification of Mandarin tones (Gauthier et al. 2007). In connected speech, Mandarin f_0 patterns display a greater amount of dynamic movement as a function of time and number of syllables than English (Eady 1982). Thus, early cortical processing of native pitch contours may be enhanced by a more efficient acoustic change-detection process that has been shaped by long-term experience with rapidly changing pitch.

Cross-Language Comparisons of Auditory Electrophysiology and Behavior

Cross-language research examining preattentive processing of linguistic stimuli often uses stimuli that are acoustically controlled so that the standard and the deviant stimuli differ from each other on only a single dimension (Näätänen et al. 1997; Sharma & Dorman 2000). However, ecologically valid stimuli typically vary on multiple dimensions. That is, a variety of cues are available for listeners, and behavioral research has shown considerable cross-language variation in the cues used to perceive these stimuli (Iverson et al. 2003).

In the perception of lexical tones, for example, listeners can use a variety of cues, including pitch onsets, offsets, direction, overall slope, average height, and location of turning point (Gandour & Harshman 1978; Gandour 1983). Native speakers of a contour-tone language (e.g., Mandarin) tend to emphasize dynamic cues such as pitch direction and slope, whereas nontone speakers tend to place more emphasis on cues related to pitch height. In the current study, ecologically valid stimuli representative of pitch contours that occur in Mandarin natural speech were used to determine electrophysiological correlates of pitch perception. The within-category stimulus pair (T1/T2i) varies on more pitch dimensions and therefore is more salient perceptually than the between-category pair (T2/T2i). Consequently, we cannot unequivocally tease apart acoustic and categorical processes underlying the enhancement of MMN in native speakers by means of MMN responses alone.

Our behavioral data, on the other hand, clearly demonstrate categorical effects on native speakers' discrimination of the within-category T1/T2i pair. Such effects are not evident during preattentive stages of processing, where both Mandarin and English speakers exhibit larger MMN responses to the acoustically salient condition (T1/T2i). Thus, we argue that a purely categorical explanation of MMN modulation in cross-language experiments does not suffice. Further experiments are required to assess more fully the relative contributions of sensory and phonetic components underlying pitch perception.

CONCLUSION

Long-term experience with a tone language enhances preattentive cortical processing of pitch contours that occur in natural speech regardless of phonetic category. This finding suggests that experience-dependent effects, as indexed by MMN responses to linguistically relevant pitch contours, predominantly reflect the malleability of acoustical encoding of pitch at early stages of auditory processing. At attentive stages of processing, however, tonal category has a profound effect on discrimination accuracy.

ACKNOWLEDGMENTS

The authors thank Jayaganesh Swaminathan for his assistance in generating the IRN stimuli and Bruce Craig and Eunjung Lim for their help with the statistical analysis.

This work was supported by NIH research grant R01 DC008549 (to A.K.), College of Liberal Arts (to A.K., J.T.G.), and New Century doctoral fellowship, ASHA Foundation (to B.C.).

Address for correspondence: Ananthanarayan Krishnan, Department of Speech Language Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038. E-mail: rkrish@purdue.edu.

Bharath Chandrasekaran is now affiliated with the Roxelyn and Richard Pepper Department of Communication Sciences, Northwestern University, USA.

Received September 7, 2008; accepted March 24, 2009.

REFERENCES

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Lang Speech*, *21*, 319–325.

Banai, K., Nicol, T., Zecker, S. G., et al. (2005). Brainstem timing: Implications for cortical processing and literacy. *J Neurosci*, *25*, 9850–9857.

Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychol Bull*, *133*, 651–672.

Chandrasekaran, B., Gandour, J. T., Krishnan, A. (2007a). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restor Neurol Neurosci*, *25*, 195–210.

Chandrasekaran, B., Krishnan, A., Gandour, J. T. (2007b). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport*, *18*, 1963–1967.

Chandrasekaran, B., Krishnan, A., Gandour, J. T. (2007c). Mismatch negativity to pitch contours is influenced by language experience. *Brain Res*, *1128*, 148–156.

Chandrasekaran, B., Krishnan, A., Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang, 108*, 1–9.

Cheour, M., Shestakova, A., Alku, P., et al. (2002). Mismatch negativity shows that 3–6-year-old children can learn to discriminate non-native speech sounds within two months. *Neurosci Lett*, *325*, 187–190.

Eady, S. J. (1982). Differences in the F_0 patterns of speech: Tone language versus stress language. *Lang Speech*, *25*, 29–42.

Fujioka, T., Trainor, L. J., Ross, B., et al. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *J Cogn Neurosci*, *16*, 1010–1021.

Fujioka, T., Trainor, L. J., Ross, B., et al. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J Cogn Neurosci*, *17*, 1578–1592.

Gandour, J. (1983). Tone perception in Far Eastern languages. *J Phonet*, *11*, 149–175.

Gandour, J., & Harshman, R. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Lang Speech*, *21*, 1–33.

Gauthier, B., Shi, R., Xu, Y. (2007). Learning phonetic categories by tracking movements. *Cognition*, *103*, 80–106.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57.

Joanisse, M. F., Robertson, E. K., Newman, R. L. (2007). Mismatch negativity reflects sensory and phonetic speech processing. *Neuroreport*, *18*, 901–905.

Kraus, N., & Cheour, M. (2000). Speech sound representation in the brain. *Audiol Neurootol*, *5*, 140–150.

Kraus, N., McGee, T., Carrell, T. D., et al. (1995). Central auditory system plasticity associated with speech discrimination training. *J Cogn Neurosci*, *7*, 25–32.

Krishnan, A., Swaminathan, J., Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci, 21*, 1092–1105.

Krishnan, A., Xu, Y., Gandour, J. T., et al. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res*, *25*, 161–168.

Kujala, T., Karma, K., Ceponiene, R., et al. (2001). Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *Proc Natl Acad Sci USA*, *98*, 10509–10514.

Kujala, T., & Näätänen, R. (2001). The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neurosci Biobehav Rev*, *25*, 535–543.

Kujala, T., Tervaniemi, M., Schroger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biol Psychol*, *74*, 1–19.

Li, P., Sepanski, S., Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behav Res Methods*, *38*, 202–210.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Luo, H., Ni, J. T., Li, Z. H., et al. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc Natl Acad Sci USA*, *103*, 19558–19563.

Maess, B., Jacobsen, T., Schroger, E., et al. (2007). Localizing pre-attentive auditory memory-based comparison: Magnetic mismatch negativity to pitch change. *Neuroimage*, *37*, 561–571.

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

Näätänen, R., Lehtokoski, A., Lennes, M., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Näätänen, R., Paavilainen, P., Rinne, T., et al. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin Neurophysiol*, *118*, 2544–2590.

Nenonen, S., Shestakova, A., Huotilainen, M., et al. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain Lang*, *92*, 26–32.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Pulvermuller, F., Shtyrov, Y., Hasting, A. S., et al. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain Lang*, *104*, 244–253.

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

Shtyrov, Y., Kujala, T., Palva, S., et al. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *Neuroimage*, *12*, 657–663.

Song, J. H., Skoe, E., Wong, P. C., et al. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci*, *20*, 1–11.

Suga, N., Gao, E., Zhang, Y., et al. (2000). The corticofugal system for hearing: Recent progress. *Proc Natl Acad Sci USA*, *97*, 11807–11814.

Suga, N., Ma, X., Gao, E., et al. (2003). Descending system and plasticity for auditory signal processing: Neuroethological data for speech scientists. *Speech Commun*, *41*, 189–200.

Suga, N., Xiao, Z., Ma, X., et al. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron*, *36*, 9–18.

Swaminathan, J., Krishnan, A., Gandour, J. T. (2008). Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Trans Biomed Eng*, *55*, 281–287.

Tervaniemi, M., Jacobsen, T., Rottger, S., et al. (2006). Selective tuning of cortical sound-feature processing by language experience. *Eur J Neurosci*, *23*, 2538–2541.

Trainor, L. J., McDonald, K. L., Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *J Cogn Neurosci*, *14*, 430–442.

Tremblay, K., Kraus, N., Carrell, T. D., et al. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *J Acoust Soc Am*, *102*, 3762–3773.

Winer, B. J., Brown, R., Michels, K. M. (1991). *Statistical Principles in Experimental Design* (3rd ed.). New York: McGraw-Hill.

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl Psycholinguist*, *28*, 565–585.

Wong, P. C., Skoe, E., Russo, N. M., et al. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci*, *10*, 420–422.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *J Phonet*, *25*, 61–83.

Xu, Y. (2006). Tones in connected discourse. In K. Brown (Ed.). *Encyclopedia of Language and Linguistics* (2nd ed., Vol. 12, pp, 742–750). Oxford, UK: Elsevier.

Yost, W. A. (1996a). Pitch of iterated rippled noise. *J Acoust Soc Am*, *100*, 511–518.

Yost, W. A. (1996b). Pitch strength of iterated rippled noise. *J Acoust Soc Am*, *100*, 3329–3335.