Published in final edited form as:

Ear Hear. 2009 October ; 30(5): 552–558. doi:10.1097/AUD.0b013e3181a7e1c2.

Sensory processing of linguistic pitch as reflected by the mismatch negativity

Bharath Chandrasekaran, PhD¹, Ananthanarayan Krishnan, PhD¹, and Jackson T. Gandour, PhD¹

¹Department of Speech Language Hearing Sciences, Purdue University, USA

Abstract

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 to 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 mismatch negativity is a useful index in examining long-term plasticity to linguistically-relevant acoustic features.

Keywords

pitch; human; iterated rippled noise (IRN); mismatch negativity (MMN); Mandarin Chinese

Introduction

The traditional mismatch negativity (MMN) is an automatic change-related fronto-centrally distributed cortical potential elicited using 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,

Author responsible for correspondence: Jackson T. Gandour, PhD, Purdue University, Department of Speech Language Hearing Sciences, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038 USA, Fax 765.494.0771; Tel 765.494.3821, Email: gandour@purdue.edu.

Author to whom requests for reprints should be sent: Ananthanarayan Krishnan (rkrish@purdue.edu)

the brain's response to 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 (Naatanen, 2001; Naatanen et al., 2007). The MMN has been extensively utilized in basic and clinical research as an objective index of central sound representation and auditory discrimination (Bishop, 2007; Kraus & Cheour, 2000; Kraus, et al., 1995; Kujala & Näätänen, 2001; Kujala, Tervaniemi, & Schroger, 2007).

The MMN has also been shown to be influenced by long-term experience and short-term training. Crosslanguage MMN experiments have demonstrated experience-dependent modulation of *segmental* information (consonants, vowels) (Naatanen, 2001; Sharma & Dorman, 2000) and *suprasegmental* units (vowel quantity, lexical tones) (Chandrasekaran, Krishnan, & Gandour, 2007b; Nenonen, Shestakova, Huottilainen, & Naatanen, 2005; Tervaniemi, et al., 2006) in speech. Short-term training paradigms have demonstrated an increased magnitude of the MMN as a function of training (Cheour, Shestakova, Alku, Ceponiene, & Näätänen, 2002; Kraus, et al., 1995; Kujala, et al., 2001; Tremblay, Kraus, Carrell, & McGee, 1997). Based on 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 (Naatanen, 2001; Naatanen, Paavilainen, Rinne, & Alho, 2007). Since 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 (Naatanen, et al., 2007). Similarly, processing advantages post-training 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 brainstem, native speakers exhibit more accurate representation of linguistically relevant pitch contours (Krishnan, Swaminathan, & Gandour, in press; Krishnan, Xu, Gandour, & Cariani, 2005). At the level of the cerebral cortex, the preattentive MMN response to segmental (Joanisse, Robertson, & Newman, 2007) and suprasegmental (Chandrasekaran, Krishnan, & Gandour, in press) 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 (Naatanen, et al., 2007). Luo et al. propose 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 (Chandrasekaran, Krishnan, & Gandour, 2007a; Chandrasekaran, Krishnan, et al., 2007b; Tervaniemi, et al., 2006). 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 crosslanguage (Chinese, English) study was to examine the extent to which acoustic and categorical processes contribute towards 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 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 vs. 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, irrespective 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 predominantly driven by acoustics ($T1/T2i > T2/T2i$). Further, by utilizing 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 are 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.

Materials and Methods

Subjects

Eleven, native speakers of Mandarin Chinese (5 male; 6 female) and eleven adult, native speakers of American English (5 male; 6 female) participated in the ERP experiment. The two groups were closely matched with respect to age (Chinese: $M = 24.4$, $SD = 4.2$; English: $M = 26.2$, $SD = 3.2$) and number of years of formal education (Chinese: $M = 17.4$, $SD = 2.3$; English: $M = 18.4$, $SD = 2.4$). 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 years 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, Sepanski, & Zhao, 2006). Native speakers of Mandarin Chinese had no English instruction before the age of 11 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.

Stimuli

Three time-varying nonspeech (iterative ripple noise, IRN) f_0 contours were created using procedures described in Swaminathan, Krishnan, and Gandour (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, 1996b). 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, are devoid of formant structure. Of the three time-varying f_0 contours (Fig. 1), two (T1, 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 f_0 contour. The third stimulus (T2i) 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, Xu 2006). The duration of all four stimuli was fixed at 250 ms including the 10 ms rise and fall ramps to reduce spectral splatter. Waveforms and narrowband spectrograms of the stimuli are displayed in Fig. 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. Participants 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 ms. Within an oddball sequence, the frequent stimulus (standard) occurred at a probability of 0.85 and the infrequent stimulus with 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 either in the context of 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 till one hundred artifact-free deviants were collected for each sequence. The experiment ran for approximately 2 hours 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 (Biologic TIP-300).

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 μ V. The electrical signal was bandpass filtered at 1–30 Hz.

Data analysis—The baseline for the grand averaged waveforms was defined as the average of the amplitude values between –100 ms and 0 ms (onset of stimuli). To obtain the MMN, the deviant waveforms from the T1/T2 and T2/T2i sequences were subtracted from standard waveforms presented, respectively, in the T2i/T1 and T2i/T1 sequences. 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 125–300 ms. The MMN mean amplitude was calculated as the mean voltage from a 100 ms window centered on the MMN peak latency. The mean amplitude measure is a more robust index relative to peak amplitude measures (Bishop, 2007; Luck, 2005). In previous studies, this index has consistently reflected experience-

dependent effects related to the MMN (Chandrasekaran, Krishnan, et al., 2007a, 2007b; Chandrasekaran, et al., in press).

The MMN mean amplitude and peak latencies were analyzed using a three-way mixed model 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, T2i) immediately following the ERP experiment. They were first presented with a practice set of stimuli in order 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 ms 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 random order. Subjects were asked to press the left (‘same’) or right (‘different’) mouse button to indicate their discrimination judgment during the 1.5 s response interval following each pair. Stimuli were presented binaurally by means of computer playback (E-Prime) through 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 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, Brown, & Michels, 1991) were subjected to a mixed model ANOVA to examine whether discrimination accuracy 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 Fig. 3. Both conditions (T1/T2i, T2/T2i) elicited robust MMN responses for both groups within the 125–300 ms 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 X condition X 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 are displayed in the left panel of Fig. 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 Fig. 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. Though 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 X tonal pair) repeated measures ANOVA conducted on the arcsine-transformed proportion of 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 Fig. 5. Per group, only the Chinese exhibited significant differences in response accuracy between pairs [$F_{1,20} = 117.02, p < 0.0001, \eta^2_{\text{partial}} = 0.85$]; cf. 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 crosslanguage study suggest that language experience modulates the automatic early cortical processing of native pitch contours. Modulation of the mismatch negativity, an index of early cortical processing, appears to be sensitive to acoustic dimensions, whereas behavioral discrimination in native speakers appears to be exclusively determined by categorical effects. Chinese speakers, as compared to English, exhibit larger MMN responses and earlier peak latencies across conditions (T1/T2i; T2/T2i). Both groups, however, show a larger MMN in response to the T1/T2i than T2/T2i condition. This is presumably due to the fact that T2/T2i varies on fewer pitch dimensions and/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 contrast (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. Since 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 which are differentially weighted by language experience (Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, et al., 2007a, 2007b). 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 to 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 (Luo, et al., 2006; Shtyrov, Kujala, Palva, Ilmoniemi, & Naatanen, 2000). Lexical tones and consonants are distinct acoustically. The former are characterized by slowly-changing spectral variation; the latter, 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 for the T1/T2i condition. Yet they are less accurate than English in distinguishing perceptually between these two pitch contours. This disparity between electrophysiology and behavior suggest 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 (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004, 2005; Trainor, McDonald, & Alain, 2002). Taken together, these data suggest a strong acoustic bias for the mismatch negativity irrespective of the nature of the stimulus, i.e. speech, non-speech, 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., in press).

As far as the MMN itself is concerned, we are unable to completely disentangle sensory vs. 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, Jacobsen, Schroger, & Friederici, 2007). The early part of the MMN (105–125 ms) is mainly due to a non-comparison, 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 analysis to a 125–300 ms 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 primarily determined 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 brainstem. For example, Mandarin speakers show more robust pitch representation than English 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 to nonmusicians (Wong, Skoe, Russo, Dees, & Kraus, 2007). By using IRN stimuli, the MMN responses are not confounded by lexical-semantic interference, which could give native speakers an advantage (Pulvermuller, Shtyrov, Hasting, & Carlyon, 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, Krishnan, et al., 2007a).

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 brainstem (Krishnan, et al., 2005; Song, Skoe, Wong, & Kraus, 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 results 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, Gao, Zhang, Ma, & Olsen, 2000; Suga, Ma, Gao, Sakai, & Chowdhury, 2003; Suga, Xiao, Ma, & Ji, 2002). The cortex improves the signal quality of its input by shaping the brainstem 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 brainstem and cortex. Abnormal brainstem timing in learning disabilities is related to a reduction in cortical sensitivity to acoustic change and deficient literacy skills (Banai, Nicol, Zecker, & Kraus, 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, Shi, & Xu, 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 used stimuli that are acoustically controlled so that the standard and the deviant stimuli only differ from each other on a single dimension (Naatanen, 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 utilized to perceive these stimuli (Iverson, et al., 2003).

In discriminating between lexical tones for example, listeners can use a variety of cues, including pitch onsets, offsets, direction, overall slope, average height, location of turning point (Gandour, 1983; Gandour & Harshman, 1978). Native speakers of a contour-tone language (e.g., Mandarin) tend to emphasize dynamic cues such as pitch direction and slope, whereas non-tone 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 more fully assess 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 irrespective of phonetic category. This finding suggests that experience-dependent effects, as indexed by MMN responses to linguistically-relevant pitch contours, predominantly reflects 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

Thanks to Jayaganesh Swaminathan for his assistance in generating the IRN stimuli; Bruce Craig and Eunjung Lim for their help with statistical analysis.

Sources of support: NIH research grant R01 DC008549 (A.K.); College of Liberal Arts (A.K.; J.T.G.); New Century doctoral fellowship, ASHA Foundation (B.C.)

Abbreviations

ERP	evoked response potential
IRN	iterated rippled noise
MMN	mismatch negativity

References

- Abramson AS. Static and dynamic acoustic cues in distinctive tones. *Language and Speech* 1978;21:319–325. [PubMed: 750791]
- Banai K, Nicol T, Zecker SG, Kraus N. Brainstem timing: implications for cortical processing and literacy. *Journal of Neuroscience* 2005;25(43):9850–9857. [PubMed: 16251432]
- Bishop DVM. Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychological Bulletin* 2007;133(4):651–672. [PubMed: 17592960]
- Chandrasekaran B, Gandour JT, Krishnan A. Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience* 2007;25(3–4):195–210. [PubMed: 17942999]
- Chandrasekaran B, Krishnan A, Gandour JT. Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport* 2007a;18(3):1963–1967. [PubMed: 18007195]
- Chandrasekaran B, Krishnan A, Gandour JT. Mismatch negativity to pitch contours is influenced by language experience. *Brain Research* 2007b;1128(1):148–156. [PubMed: 17125749]

- Chandrasekaran B, Krishnan A, Gandour JT. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*. in press.
- Cheour M, Shestakova A, Alku P, Ceponiene R, Näätänen R. Mismatch negativity shows that 3-6-year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters* 2002;325(3):187–190. [PubMed: 12044652]
- Eady SJ. Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech* 1982;25(1):29–42.
- Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C. Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience* 2004;16(6):1010–1021. [PubMed: 15298788]
- Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C. Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience* 2005;17(10):1578–1592. [PubMed: 16269098]
- Gandour J. Tone perception in Far Eastern languages. *Journal of Phonetics* 1983;11:149–175.
- Gandour J, Harshman R. Crosslanguage differences in tone perception: a multidimensional scaling investigation. *Language and Speech* 1978;21:1–33. [PubMed: 692240]
- Gauthier B, Shi R, Xu Y. Learning phonetic categories by tracking movements. *Cognition* 2007;103(1):80–106. [PubMed: 16650399]
- Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, et al. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 2003;87(1):B47–B57. [PubMed: 12499111]
- Joanisse MF, Robertson EK, Newman RL. Mismatch negativity reflects sensory and phonetic speech processing. *Neuroreport* 2007;18(9):901–905. [PubMed: 17515798]
- Kraus N, Cheour M. Speech sound representation in the brain. *Audiol Neurotol* 2000;5(3–4):140–150. [PubMed: 10859409]
- Kraus N, McGee T, Carrell TD, King C, Tremblay K, Nicol T. Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience* 1995;7(1):25–32.
- Krishnan A, Swaminathan J, Gandour JT. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*. in press.
- Krishnan A, Xu Y, Gandour JT, Cariani P. Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Research. Cognitive Brain Research* 2005;25(1):161–168. [PubMed: 15935624]
- Kujala T, Karma K, Ceponiene R, Belitz S, Turkila P, Tervaniemi M, et al. Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *Proceedings of the National Academy of Sciences of the United States of America* 2001;98(18):10509–10514. [PubMed: 11517333]
- Kujala T, Näätänen R. The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neuroscience & Biobehavioral Reviews* 2001;25(6):535–543. [PubMed: 11595273]
- Kujala T, Tervaniemi M, Schroger E. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol Psychol* 2007;74(1):1–19. [PubMed: 16844278]
- Li P, Sepanski S, Zhao X. Language history questionnaire: A web-based interface for bilingual research. *Behavioral Research Methods* 2006;38(2):202–210.
- Luck, SJ. An introduction to the event-related potential technique. Cambridge, MA: MIT Press; 2005.
- Luo H, Ni JT, Li ZH, Li XO, Zhang DR, Zeng FG, et al. Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103(51):19558–19563. [PubMed: 17159136]
- Maess B, Jacobsen T, Schroger E, Friederici AD. Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *Neuroimage* 2007;37(2):561–571. [PubMed: 17596966]
- Naatanen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 2001;38(1):1–21. [PubMed: 11321610]

- Naatanen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 1997;385(6615):432–434. [PubMed: 9009189]
- Naatanen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology* 2007;118(12):2544–2590. [PubMed: 17931964]
- Neenonen S, Shestakova A, Huotilainen M, Naatanen R. Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language* 2005;92(1):26–32. [PubMed: 15582033]
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97–113. [PubMed: 5146491]
- Pulvermuller F, Shtyrov Y, Hasting AS, Carlyon RP. Syntax as a reflex: neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language* 2008;104(3):244–253. [PubMed: 17624417]
- Sharma A, Dorman MF. Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America* 2000;107(5):2697–2703. [PubMed: 10830391]
- Shtyrov Y, Kujala T, Palva S, Ilmoniemi RJ, Naatanen R. Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *Neuroimage* 2000;12(6):657–663. [PubMed: 11112397]
- Song JH, Skoe E, Wong PC, Kraus N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience* 2008;20(10):1–11. [PubMed: 17919082]
- Suga N, Gao E, Zhang Y, Ma X, Olsen JF. The corticofugal system for hearing: recent progress. *Proceedings of the National Academy of Sciences of the United States of America* 2000;97(22):11807–11814. [PubMed: 11050213]
- Suga N, Ma X, Gao E, Sakai M, Chowdhury SA. Descending system and plasticity for auditory signal processing: neuroethological data for speech scientists. *Speech Communication* 2003;41:189–200.
- Suga N, Xiao Z, Ma X, Ji W. Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 2002;36(1):9–18. [PubMed: 12367501]
- Swaminathan J, Krishnan A, Gandour JT. Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Trans Biomed Eng* 2008;55(1):281–287. [PubMed: 18232372]
- Tervaniemi M, Jacobsen T, Rottger S, Kujala T, Widmann A, Vainio M, et al. Selective tuning of cortical sound-feature processing by language experience. *European Journal of Neuroscience* 2006;23(9):2538–2541. [PubMed: 16706861]
- Trainor LJ, McDonald KL, Alain C. Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience* 2002;14(3):430–442. [PubMed: 11970802]
- Tremblay K, Kraus N, Carrell TD, McGee T. Central auditory system plasticity: generalization to novel stimuli following listening training. *J Acoust Soc Am* 1997;102(6):3762–3773. [PubMed: 9407668]
- Winer, BJ.; Brown, R.; Michels, KM. Statistical principles in experimental design. Vol. 3rd ed. New York: McGraw-Hill; 1991.
- Wong PC, Perrachione TK. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 2007;28(4):565–585.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 2007;10(4):420–422.
- Xu Y. Contextual tonal variations in Mandarin. *Journal of Phonetics* 1997;25:61–83.
- Xu, Y. Tones in connected discourse. In: Brown, K., editor. *Encyclopedia of Language and Linguistics*. Vol. 2nd ed. Vol. 12. Oxford, UK: Elsevier; 2006. p. 742–750.
- Yost WA. Pitch of iterated rippled noise. *Journal of the Acoustical Society of America* 1996a;100(1):511–518. [PubMed: 8675844]
- Yost WA. Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America* 1996b;100(5):3329–3335. [PubMed: 8914314]

Figure 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 a 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 the standard in one condition (T1/T2i); 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 high 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.

Figure 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 broad band 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 for amplitude to be controlled independent of pitch contour. IRN, iterated rippled noise.

Figure 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). Irrespective 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 irrespective of condition.

Figure 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. Irrespective of condition, the MMN peaked earlier for the Chinese group relative to the English group.

Figure 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.