Preattentive Auditory Processing of Pitch Glide

Feng Gu 1, Xiao-Dong Wang 2, Kang He 2, Lin Chen 2

1 CAS Key Laboratory of Brain Function and Diseases, School of Life Sciences, University of Science and Technology of China, Hefei, China
2 Auditory Research Laboratory, University of Science and Technology of China, Hefei, China

guark@mail.ustc.edu.cn, linchen@ustc.edu.cn

Abstract

Pitch glide is a continuously rising or falling characteristic of pitch contour in tonal languages, such as Mandarin Chinese, and is critical for tonal speech comprehension. An interesting but unanswered question is whether pitch glide can be automatically processed at a preattentive stage in the brain. To address this question, we used iterated rippled noise (IRN) as a linguistically meaningless nonspeech stimulus to evoke mismatch negativity (MMN), an indicator of automatic auditory processing, from native speakers of Mandarin Chinese. To create a pitch glide contrast, we frequently presented IRN to the subjects with a flat pitch contour at different pitch levels and infrequently varied its pitch contour from flat to flat-rise. The pitch glide contrast evoked a robust MMN response with the peak latency around 155 ms after the rising onset. In addition, the steeper rising slope was, the larger MMN was evoked. These results suggest a preattentive auditory analysis of pitch glide information, which may be fundamental for contour tone perception.

Index Terms: pitch glide, lexical tone, mismatch negativity, preattentive auditory perception, Mandarin Chinese

1. Introduction

About 60-70% of the world’s languages are tonal [1], which use tones to signal word meaning besides consonants and vowels. In Mandarin Chinese, a tonal language, a great number of words differ from each other solely in lexical tone. The primary acoustic cue for tone perception is the fundamental frequency (f0) contour of the voice. In other words, listeners mainly depend on f0 contour, which is perceived as a pitch contour, to distinguish lexical tones [2-4]. F0 contour has two characteristics [5, 6]: f0 level (f0 height), which gives rise to the perceived pitch level such as high or low, and f0 glide (f0 slope), which gives rise to the perceived pitch glide such as rising or falling. In the register tone system, such as most Bantu languages, the distinguishing feature of tones is the relative pitch level (high, mid, or low). These tones have a flat (static) f0 contour and are called level tone (register tone). In contrast, in the contour tone system such as Mandarin Chinese language, the distinguishing feature of tones is the pitch glide (falling or tone 4; rising or tone 2; dipping or tone 3) (Figure 1). These tones with non-flat f0 contour are called contour tones. To perceive a level tone in a register tone system, the speech perception system needs to analyze its relative pitch level (rather than absolute pitch level) whereas to perceive a contour tone, the system has to use the pitch glide information as a primary cue [7, 8]. Thus, pitch glide processing is essential to contour tone language perception.

Preattentive auditory processing of contour tone, which can be indexed with mismatch negativity (MMN), has been described in several studies using speech and/or non-speech stimuli [5, 6, 10-17]. In these studies, MMN was successfully elicited with oddball paradigm in which one type of pitch contour served as standard, and another type of pitch contour served as deviant. However, it is not clear from these studies whether the evoked MMN was contributed by pitch level or pitch glide, or by both. Since previous studies showed that pitch level deviation could elicit an MMN response [18-21], there remains a question whether the pitch glide could also elicit an MMN response. In other words, there remains a question whether pitch glide information can be automatically analyzed in the brain at a preattentive stage. To answer this question, we used iterated rippled noise (IRN) as a linguistically meaningless nonspeech stimulus for evoking MMN from native speakers of Mandarin Chinese. We found that the pitch glide contrast can evoke a robust MMN response, indicating a preattentive analysis of pitch glide information in the brain.

2. Methods

2.1. Participants

Twelve native speakers of Mandarin Chinese (seven males and five females, 18-24 years old, right-handed, and musically untrained) with normal hearing participated in the study. An informed written consent was obtained from each subject.
2.2. Stimuli
IRN stimuli were generated with f0 contours using procedures similar to those described in a previous study by others [22]. A high iteration step (32) was used for all stimuli with a gain set to 1. The IRN stimuli showed clear energy bands at the f0 and its integer multiples without formant structure or modulated temporal envelope (Figure 2A). Seven IRN stimuli (F1-F7) with flat f0 contour were generated and served as abstract standards. Four IRN stimuli (R1-R4) with flat-rise f0 contour were generated and served as deviants. R1, R2 and R3 possessed flat f0 contour during the initial 100 ms, and were followed by a rising f0 glide with different slope (R1: 10 Hz/100 ms; R2: 30 Hz/100 ms; R3: 50 Hz/100 ms). R2 and R4 possessed equal rising slope, but R2 started rising at 100 ms whereas R4 started rising at 50 ms. Each stimulus has the same duration of 200 ms with 10 ms linear rise and fall time. Figure 2B shows the f0 contours of all the 11 stimuli.

![Figure 2: Narrowband spectrograms (A) and F0 contours (B) of IRN stimuli. In the oddball paradigm, F1 to F7 served as sub-standards and R1 to R4 served as sub-deviants.](image)

2.3. Procedures
The experimental stimulation paradigm consisted of an oddball block and a control block. In both blocks the subject was instructed to ignore the auditory stimulation and watch a self-selected silent movie with subtitles. Each stimulus was presented diotically at 60 dB SPL through headphones (TDH-39; Telephonics, Farmingdale, NY) with a stimulus onset asynchrony (SOA) of 700 ms in an electrically shielded and soundproof room.

In the oddball block, an “abstract standards oddball paradigm” was used. The paradigm was constructed as follows: a set of eight digits (from 1 to 8) were multiply presented in a sequence, and in each set the eight digits were presented in a random order. Then, the overall sequence was slightly modified to make the adjacent same digits were divided by at least two different digits. Afterwards, the digits 1 to 7 were replaced by corresponding sub-standard (F1-F7), and digit 8 was replaced by a random sub-deviant (R1-R4). In such an oddball paradigm, each sub-standard had a probability of 12.5% and each sub-deviant had a probability of 3.125%. At least two sub-standards were presented between any two adjacent sub-deviants. The oddball block was divided into four sequences with 5 min break between each other, and in each sequence the initial 10 stimuli were always sub-standards and a total of 1450 stimuli were delivered. Overall, each type of sub-deviant was delivered about 180 trials in the oddball block.

In the control block, there were four sequences with 5 min break between each other, and in each sequence one type of sub-deviants (R1-R4) was delivered in 400 trials and served as control standard of that sub-deviant. Control standard and sub-deviant were designed to be physically identical. The MMN response was derived by subtract the event related potential (ERP) of control standard from that of sub-deviant.

The control block was carried out 10 min after oddball block. Within the control block, the order of sequence was counterbalanced among subjects. The experiment ran for approximately 2.5 hr for each subject including electrode application and removal.

2.4. EEG recordings
Whole-head recordings of EEG were obtained from each subject using SynAmps amplifier (Neuroscan, Sterling, VA) with a cap carrying 64 Ag/AgCl electrodes placed on the scalp at locations following the 10% system [23]. In addition, electrical activities were recorded from both mastoids (left: LM and right: RM). The reference electrode was attached to the tip of nose, horizontal EOG was recorded using bipolar channel placed lateral to the outer canthi of the two eyes and vertical EOG was recorded using bipolar channel placed above and below the left eye. Skin impedances were kept <5kΩ. Alternating current signals (0.5-100 Hz) were continuously recorded and digitized with a 16-bit resolution at a sampling rate of 500 Hz.

2.5. Data analysis
The data from the whole-head recordings were first processed to correct EOG artifacts using the method described in a previous study [24]. Then, the EEG data was off-line band-pass filtered (1-25 Hz) with a finite impulse response (FIR) filter. Epochs were set as 600 ms long, including a 100 ms pre-stimulus baseline. Epochs with amplitudes exceeding ±50 μV at any channel except EOG channels were excluded from averaging. ERPs were averaged for each sub-deviant (R1-R4) and its control standard and were corrected relative to the
baseline (100 ms pre-stimulus mean amplitude). For each sub-deviant, MMN was derived by subtracting the ERP of its control standard from that of sub-deviant itself. For statistical analysis, location Fz was taken as the region of interest because MMN was known to show a fronto-central distribution and most prominent at the fronto-central sites [25, 26]. For each subject, the MMN peak latency was the time when most negative amplitude occurred within a time window between 150 and 300 ms from stimulus onset, and the MMN amplitude was calculated as the mean voltage from a 40 ms window centered on the grand averaged MMN peak. Independent sample t tests were performed to determine whether the MMN amplitudes significantly differed from zero. Statistical analysis of MMN amplitude was performed using one-way repeated measures analysis of variance (RM-ANOVA) with sub-deviant type (R1, R2 and R3, each denotes different magnitude of rising slope, i.e. R1 < R2 < R3) as the factor. Statistical analysis of MMN peak latency was conducted using one way RM-ANOVA with sub-deviant type (R2 and R4, each denotes different onset of rising) as the factor. Greenhouse-Geisser (G-G) corrections of degrees of freedom were made when appropriate and G-G epsilon values were reported. Bonferroni adjustments were made in post hoc multiple comparison procedures.

3. Results

3.1. MMN responses were elicited by sub-deviants

Each sub-deviant type (R1-R4) elicited a robust MMN response within the 100-300 ms time window from stimulus onset. The grand averaged MMN waveforms at 5 typical locations (LM, F3, Fz, F4 and RM) were shown in Figure 3A. Each MMN response reversed in polarity at mastoid locations (LM, RM). The topographies of grand averaged MMN at peak latency showed a frontal-central distribution (Figure 3B). Each MMN response reversed in polarity at mastoid locations (LM, F3, Fz, F4 and RM) were shown in Figure 3A. The topographies of grand averaged MMN at peak latency showed a frontal-central distribution (Figure 3B). These results were consistent with the opinion that MMN peak latency decreases as the magnitude of deviation increases till a plateau is reached [27-29].

We calculated the peak latencies of R1-, R2-, R3- and R4-MMN again by measuring the time point when most negative peak occurred from rising onset in sub-deviant (i.e. the deviation onset), rather than from stimulus onset. We obtained the mean MMN peak latency of 167.3 ± 5.1 ms for R1-MMN, 151.0 ± 4.5 ms for R2-MMN, 151.3 ± 3.1 ms for R3-MMN and 150.3 ± 4.8 ms for R4-MMN (n = 12). The latency was consistent with that of typical MMN response, which peaks around 150-200 ms after deviation onset [25, 26].

3.2. The magnitude of MMN increased as rising slope increased

One way RM-ANOVA with sub-deviant type (R1, R2 and R3) as the factor indicated that the MMN amplitude was significantly influenced by sub-deviant type (F(2, 22) = 17.424, ε = .666, p < .001). Post hoc test showed that R1-MMN was smaller than R2- (p = .002) and R3-MMN (p = .002), and R2-MMN was smaller than R3-MMN (p = .034), as shown in Figure 4A. In other words, the MMN amplitude increased as the rising slope (i.e. the magnitude of deviation in the oddball sequence) in sub-deviant increased. There was no significant difference in the amplitude between R2-MMN and R4-MMN (paired-Student t test, t(11) = 1.132, p = .282), which was consistent with the fact that they had the same rising slope.

3.3. MMN peak latency paralleled with rising onset time

One-way RM-ANOVA with sub-deviant type (R2 and R4) as the factor indicated that the MMN peak latency was significantly influenced by sub-deviant type (F(1, 22) = 67.736, ε = 1.000, p < .001) (Figure 4B). The mean peak latency difference between R2- and R4-MMN was 50.7 ± 6.2 ms (SEM, n = 12), which was consistent with the difference in acoustic rising onset time between R2 and R4 (i.e. 50 ms).

One-way RM-ANOVA with sub-deviant type (R1, R2 and R3) as the factor indicated that the MMN peak latency was significantly influenced by sub-deviant type (F(2, 22) = 7.846, ε = .961, p = .003). Post hoc test showed no significant difference in latency between R2- and R3-MMN (p = 1.000), whereas the latency of R1-MMN was significantly longer than those of R2- (p = .019) and R3-MMN (p = .026) (Figure 4B). These results were consistent with the opinion that MMN peak latency decreases as the magnitude of deviation increases till a plateau is reached [27-29].

We calculated the peak latencies of R1-, R2-, R3- and R4-MMN again by measuring the time point when most negative peak occurred from rising onset in sub-deviant (i.e. the deviation onset), rather than from stimulus onset. We obtained the mean MMN peak latency of 167.3 ± 5.1 ms for R1-MMN, 151.0 ± 4.5 ms for R2-MMN, 151.3 ± 3.1 ms for R3-MMN and 150.3 ± 4.8 ms for R4-MMN (n = 12). The latency was consistent with that of typical MMN response, which peaks around 150-200 ms after deviation onset [25, 26].

Figure 3: (A) Grand-average MMN responses (n=12) elicited by deviants (R1-R4) at locations: LM, F3, Fz, F4, and RM. (B) Scalp topographic maps constructed from grand average MMN responses elicited by sub-deviants (R1-R4).
en}). We predict that the MMN evoked by pitch glide deviation would be larger in contour tone language speakers than non-tonal language speakers. This is an interesting topic deserving further study.

4.2. Pitch level elicited MMN versus pitch glide elicited MMN

In previous contour tone studies [5, 6, 10-14, 16, 17], MMN was elicited using one type of f0 contour as standard and another type as deviant. Thus, there was a co-occurrence of f0 level deviation and f0 glide deviation. Some studies show that f0 level deviation could elicit an MMN response [18-21]. In the present study, we demonstrate that f0 glide deviation could also elicit MMN response. Therefore, the two MMN responses may be additive in these studies, given that the MMN amplitude shows additivity when a deviant deviates from the standard in two or several attributes [47-51]. Nevertheless, whether or not the highly related first-order feature (f0 level) and second-order feature (f0 glide, time-varying f0 level) are additive requires further investigation.

Each tone in Mandarin Chinese has an initial flat f0 that may be a result of speech production (e.g., when the duration is 250 ms, tone 2 has an initial about 100 ms flat f0, followed by a distinct rising f0 glide, Figure 1). In the classical oddball paradigm using one type of lexical tone as standard and another type as deviant, the preceding onset f0 level deviation always precedes the following f0 glide deviation. Some studies show that, for two temporally separate deviations from the same standard, only the first one elicits an MMN if these two deviations occur within the temporal window of integration (TWI, about 150 ms, 52-55, for a review, see [25]). Thus, in the previous studies using Mandarin Chinese lexical tone as stimuli, the f0 glide deviation elicited MMN might be masked by the first f0 level deviation. This notion is consistent with the findings from former contour tone studies that only one MMN was always elicited in the latency to the preceding pitch level deviation onset. Because f0 glide information provides primary cue for contour tone perception, the language-specific MMN subcomponent that reflecting the activation of memory traces of sound cues might be missed in these studies. However, on the other hand, under circumstances that one MMN is elicited by temporally separate two deviations (double deviations) [55], the MMN is larger than that elicited by the first deviation alone, indicating that the information in second deviation might not be missed.

5. Conclusions

We demonstrate that pitch glide can evoke a robust MMN response from native speakers of Mandarin Chinese, a tonal language, indicating automatic processing of pitch glide in the analyzed. The early auditory processing of f0 glide information may be fundamental for contour tone perception.

In the present study, all the subjects were native Mandarin Chinese speakers, who analyze the pitch glide information in every day communications. Luo et al [37] found that Chinese subjects perform better than English subjects in the task of identifying pitch glide in non-speech FM (frequency-modulated) sweeps, suggesting contour tone language speakers are more skillful for acoustic pitch glide processing than non-tonal language speakers. Experience could modulate the MMN magnitude, presumably due to an elaborated mechanism of information acquirement underlying the generation of MMN [38-43] or due to activation of the long-term memory trace for speech units [44-46]. We predict that the MMN evoked by pitch glide deviation would be larger in contour tone language speakers than non-tonal language speakers. This is an interesting topic deserving further study.

4. Discussion

4.1. Pitch glide information is automatically processed at a preattentive stage

In the present study, we demonstrate that MMN responses can be elicited by pitch glide. The MMN amplitude and MMN peak latency indicate that the response is exactly elicited by f0 glide deviations. According to the memory-dependent comparison mechanism of MMN [25, 26, 30-34], the MMN elicited by f0 glide deviation reveals that the “flat” f0 contour of sub-standards is extracted and stored in the sensory memory. The “rising” f0 glide of deviant is analyzed and its incongruence with memory representation is detected automatically in preattentive auditory processing.

In both tonal and non-tonal languages, rapid phoneme perception is a prerequisite for reliable sound perception. Jacobsen found that F1/F2 formant information, crucial for vowel perception, is preattentively analyzed for both speech and non-speech stimuli [35, 36]. For tone languages, rapid toneme perception is an additional prerequisite for reliable phoneme comprehension. We now found that f0 glide information in non-speech stimuli is also preattentively
brain at a preattentive stage. The automatic processing of pitch glide may be fundamental for contour tone perception during tonal speech communication.

6. Acknowledgements

We thank Jayaganesh Swaminathan, Department of Speech Language and Hearing Sciences, Purdue University, for assistance of IRN stimuli generation. We also thank Yi Xu, Department of Speech, Hearing and Phonetic Sciences, University College London, for helpful discussion. This work was supported by the National Basic Research Program of China (Grant 2011CB504506) and the National Natural Science Foundation of China (Grants 30730041 and 30970977).

7. References


