The effect of musicality on the development of Mandarin prosody

Nari Rhee\textsuperscript{1}, Jianjing Kuang\textsuperscript{1}, Aoju Chen\textsuperscript{2}

\textsuperscript{1}University of Pennsylvania
\textsuperscript{2}Utrecht University
nrhee@sas.upenn.edu, kuangj@ling.upenn.edu, aoju.chen@uu.nl

Abstract

Past work has shown a link between children’s musicality and language learning. But research is still sparse on the effect of musicality on the development of prosody, which uses tonal and temporal cues also relevant for processing music. In particular, the questions of when and how musicality affects the development of various aspects of the prosodic grammar remain largely unknown. In this study, we investigate the effect of musicality on the development of focus-marking in Mandarin-speaking 4- to 6-year-olds using speech data elicited in a controlled but interactive setting. We have found that the development of focus-marking in Mandarin is only weakly affected by the learner’s musicality. Specifically, children produce adult-like distinctions between on-focus and pre-focus positions, regardless of musicality. A musicality effect is observed in the contrast between on-focus and post-focus positions only in the 4-year-olds. The limited musicality effect on focus-marking is in contrast with our previous work, in which we found that musicality has a salient effect on the lexical tone production by children younger than 6 years. Together, the current results suggest that musicality advantage in the development of prosody depends on aspects of the prosodic grammar and the stage of development.

Index Terms: musicality, focus, prosody, L1 acquisition, Mandarin

1. Introduction

There is by now mounting evidence that music and language are closely connected. Some theories assume shared neural networks for language and music processing (e.g. the OPERA hypothesis)\cite{1}. Others suggest parallels between musical structure and syntactic structure of language\cite{2,3} and between musical structure and the structure of speech prosody\cite{4}. Pertinent to the current study are the parallels between music and prosody. According to\cite{4}, both music and prosody include phrases across different time scales with phrases on a lower time scale nested inside phrases on a higher time scale (e.g. notes, beats, measures, phrases in music; feet, prosodic words, phonological phrases and intonational phrases in prosody). Each phrase has units differing in prominence; there may be restrictions on which segments can occur in a phrase. Both musical phrases and major prosodic phrases are characterized with phrase-final lengthening. Related to the structural similarities between music and language, musicality (i.e. an individual’s cognitive capability for music, either shaped by both innate aptitude or musical training) has been shown to have a positive effect on various aspects of language development\cite{5}. For example, musical training can improve perception of vowel duration, voice onset time\cite{6}, phonological awareness\cite{7} and syntactic processing in typically developing children\cite{8}, and reading skills\cite{9} and grammatical development in both typically developing children with Developmental Language Disorder or dyslexia\cite{10}.

However, research is still sparse on the effect of musicality on prosodic development in spite of the parallels between musical structure and prosodic structure and the known benefits of musicality (as a result of training) in detecting subtle changes in pitch\cite{11,12,13} and timing\cite{14,15}. Past work is primarily concerned with the effect of musical training on perception of variation in pitch height and perception of emotion via pitch cues. For example, musical training can lead to improved detection of subtle pitch variations in sentence-final position in Portuguese-speaking 8-year-olds\cite{9} and more accurate identification of emotions (i.e. anger and fear) in English-speaking 6-year-olds\cite{16}. Little is still known on how musicality influences different aspects of prosodic development.

The present study aims to address this question by investigating the effect of musical aptitude on children’s ability to mark narrow focus prosodically in Mandarin Chinese, compared to its effect on children’s tonal production. Focus marking and tone production are two different but related aspects of prosodic grammar of Mandarin Chinese. Adult-like use of prosody in focus marking requires that children can discern the need to highlight new or contrastive information to their interlocutor in addition to executing changes in prosodic parameters in the focal word and surrounding words. There is, however, no evidence that musicality makes a speaker more attentive to the pragmatic need of the interlocutor. On the other hand, prosodic marking of narrow focus in Mandarin Chinese entails the ability to manipulate pitch and duration without altering the phonological identity of the tones, which is primarily defined by pitch characteristics. The ability to produce tones contrastively in the acoustic space may thus be closely related to prosodic focus marking. Recent research on Mandarin-speaking children’s production of lexical tones has shown that children with higher musical aptitude produce tones with higher contrastivity at the age of 4 and 5 years but not at the age of 6 years\cite{17}. This result suggests that musicality promotes development of tone production in earlier stages of prosodic development. We thus hypothesise that musical aptitude will also facilitate prosodic marking of narrow focus in children under the age of 6 years.

2. Methods

2.1. Procedure

2.1.1. Participants

Forty-three typically-developing Mandarin-speaking children between the ages of 4 and 6 years (range: 4:0 to 6:11) were recruited from a kindergarten and a primary school in Beijing. Each subject participated in first a semi-spontaneous speech elicitation session and then a musicality testing session. The two sessions were run on two different days. In addition, to compare the children’s focus production to the adults’, 11 na-
ative Mandarin adult speakers’ speech elicitation data from [18] were included.

2.1.2. Speech elicitation

The speech elicitation session was conducted in an interactive picture-matching game setting [19], where the participants were led to answer the experimenter’s questions in full subject-verb-object (SVO) sentences (e.g. [1][8][5][3][5].v. 3[8][5].o. “The little bear throws the ball”). During the interaction, the experimenter used different wh-questions to elicit various focus conditions in the participants’ responses (see [20]). For example, subject wh-questions (e.g. Who throws the ball?) were used to induce a narrow focus on the object position in the participants’ responses, object wh-questions (e.g. What did the bear throw?) for a narrow focus on the verb. A contrastive focus on the verb and broad focus were also elicited, resulting in a total of 80 (if the child participated in half of the session) to 160 target sentences per participant, with tones and focus conditions controlled. However, in this study, we limited our analysis to the production of the sentence-medial monosyllabic verbs in narrowly-focused vs. non-focused conditions, using a subset of the data (60% of the elicited sentences) with a narrow focus on the verb (the verb is on-focus; focused), on the object (the verb is pre-focus; non-focused), or on the subject (the verb is post-focus; non-focused).

2.1.3. Musicality

In the musicality testing session, the participants took the tonal subset of the Primary Measures of Music Audiation (PMMA), a test designed to measure the musical aptitude of children in kindergarten to Grade 3 [21]. The participants listened to two short melodies and indicated whether they heard two identical melodies or two different ones. Two example questions were given before starting the test. Due to the short attention spans of younger children, the 4-year-olds were given 20 questions, which took about 10 minutes. Older children finished all 40 questions, taking approximately 20 minutes. Each participant’s musicality score (in percentage) was calculated by dividing the number of correct responses by the total number of questions given.

2.2. Analysis

2.2.1. Age and musicality groups

To assess how musicality affected the focus-marking abilities of the children, the focus production data from the children were divided into age and musicality groups (i.e. 4-L by the participant’s age (4, 5, or 6 years) and performance on the musicality test (L for low if below the median score of the speaker’s age group; H for high if median or above).

2.2.2. Tonal dispersion as a marker for focus

Generally, languages with prosodic focus-marking have more prominent acoustic cues in focused words than in non-focal words. In English, for instance, focused elements are produced with greater intensity, longer duration, and higher F0 (e.g. [22]). However, in the case of tonal languages such as Mandarin, because the acoustic cues used to signal focus-markings overlap with the set of cues for lexical tones, the realization of focus and tones exhibit close interactions. In Mandarin, the high level tone (Tone 1) has F0 raising in focal conditions, the mid-rising (Tone 2) and falling (Tone 4) tones involve a steeper and greater F0 change, and the low-dipping tone (Tone 3) has even lower F0 [23]. These findings indicate that the tones of words in focus are produced with more distinctive F0 height and contours, and are thus more dispersed in the overall tonal space. This study uses this notion of increased tonal dispersion in focused elements as a more comprehensive measure of focus-marking in tonal languages. The steps we have taken for quantifying the tonal dispersion of the verb when focused or non-focused are outlined in the next subsection (Section 2.2.3).

2.2.3. Quantifying the tonal dispersion

First, F0 values (STRAIGHT [24]) were extracted from 9 time-normalized subsegments of the sentence-medial monosyllabic verbs in the SVO sentences, using VoiceSauce [25]. The extracted F0 values were then z-score normalized by speaker and recording session to minimize the inter-speaker differences. Due to the interactive and spontaneous nature of the task, tokens with loud background noises, wrong target words, or octave jumps in F0 tracking were systematically and manually identified and removed. Using the cleaned F0 values from the 9 time-normalized subsegments of the syllables, we calculated how far the tonal tokens were from each other in the F0 space by calculating the average pairwise Euclidean distance from each token to all other tokens of different tonal categories (between-tone Euclidean distances). F0 values from all 9 timepoints were used in the analysis in order to maximally preserve the contour and temporal information. The calculated between-tone distances were compared across the on-focus and non-focus (pre-focus or post-focus) conditions, for each age and musicality group. Non-parametric Kruskal-Wallis and Wilcoxon rank-sum tests were performed for statistical validation of the differences across focus conditions. The p-values were adjusted for multiple comparisons using Holm’s method.

The number of participants in each age group, the number of analyzed tokens after cleaning, and the median musicality score of each age group are summarized in the Table 1.1

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Speakers</td>
<td>10</td>
<td>13</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>Number of Tokens</td>
<td>338</td>
<td>438</td>
<td>1,137</td>
<td>773</td>
</tr>
<tr>
<td>Median Musicality (%)</td>
<td>72.50</td>
<td>82.50</td>
<td>88.75</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 1: The number of participants, the number of analyzed tonal tokens, and the median musicality score in each age group

3. Results

Comparisons of the tonal dispersion (i.e. between-tone Euclidean distance in the F0 space) in narrowly-focused and non-focused conditions (pre-focus and post-focus) for each age and musicality group are summarized in Figure 1. The children’s production of the focus conditions is shown in the top three rows of panels (4, 5, and 6 years), with the lower musicality groups on the left and the higher musicality on the right column of panels. The bottom-left panel shows the results for the adults’ focus production. Overall, all age and musicality groups share a common trend. As can be seen in the box-plots of all panels, tones are more dispersed in focused (green) conditions than in pre-focus (red) or post-focus (blue) conditions. Kruskal-Wallis tests

1Balancing the number of participants in each age group (via random sampling) did not change the results.
confirm that for all age and musicality groups, there is a significant difference among the focus conditions (4-L: $p = 0.014$; 4-H, 5-L, 5-H, 6-L, 6-H, Adults: $p < 0.001$). The p-values of Wilcoxon tests with Holm’s adjustments for multiple comparisons (shown on top of the boxes in the box-plots) confirm that all children between the age of 4 and 6 years indeed have a significantly more dispersed tonal space in narrowly-focused conditions than in pre-focused conditions, regardless of the musicality (4-L: $p = 0.004$; 4-H: $p = 0.002$; 5-L: $p = 0.003$; 5-H, 6-L, 6-H, Adults: $p < 0.001$). Comparing the focused and post-focus conditions, the tonal dispersion is significantly larger in focused conditions in all age and musicality groups except the 4-year-old low musicality group (4-L: $p = 0.296$; 4-H: $p = 0.005$; 5-L, 5-H, 6-L, 6-H, Adults: $p < 0.001$).

These results suggest that higher musicality has no apparent advantage in learning to produce the focus and pre-focus conditions differently for children aged 4 to 6 years because the prosodic marking of focus in contrast to pre-focus has already been mastered at the age of 4 years. However, only 4-year-olds with higher musicality produce less dispersed tones in post-focal conditions than in focused conditions, suggesting that musicality may still play some role in the development of focus-marking in contrast to post-focal conditions, at the age of 4 years. In short, musicality has a limited effect on the development of tonal dispersion to mark focus in children younger than 5 years.

4. Discussion

In this study, we have investigated the effect of musicality on children’s prosodic development of focus-marking in Mandarin. Especially given that a strong musicality effect has been observed in Mandarin tone development under the age of 6 years [17] and the relevance of tonal production ability for prosodic focus marking in Mandarin, we have hypothesized that higher musicality will transfer to an earlier development of another aspect of prosodic grammar: focus-marking.

A brute-force estimation of the between-tone Euclidean distances has been taken as a proxy for modelling the tonal dispersion phenomena in prosodic focus, where the F0 height and contour differences among the tones are maximized when focused. Contrary to the hypothesis, the results reveal a rather weak influence of musicality on children’s focus-marking abilities, unlike its robust influence on tone development. An adult-like tonal dispersion difference between focused and non-focused conditions is reliably produced by almost all children regardless of musicality, and only a limited musicality effect is observed in the contrast between focused and post-focus conditions at the age of 4 years. Even in the 4-year-old low musicality group, focused words’ tones are slightly more dispersed than words in the post-focus condition, though not significantly so. In sum, the strong musicality effect found in tone development is not found in prosodic focus development.

One possible explanation for the limited musicality effect on prosodic focus marking in Mandarin-speaking children is that for younger children, producing tones contrastively may be prioritized over making the focal word more prominent and distinct from non-focal words. The ability to mark focus has to do with the children’s ability to produce contrastive tones in an efficient manner. If the F0 space is maximally used up for manifesting the tonal contrast, there is no room left for focused words to expand in space. By analogy with vowel production, if certain speakers already produce the vowels very clearly with large differences in F1 and F2 between vowels, it would...
be harder for them to introduce yet more differences in the production in formal speech. At the age of 4 years, children have incomplete tone development [18, 26, 27]. Since tones hold lexical information, these children may focus more on making a better tonal contrast, regardless of whether the word is focused or post-focal. Future investigations of the relationship between musicality and prosodic development are needed to shed further light on the role of musicality in the interplay between multiple levels of prosody that share a common set of acoustic cues.

Our results also show an asymmetric effect of musicality. That is, musicality has a weak effect on children’s ability to distinguish a focused word from its post-focal counterpart but not from its pre-focal counterpart in the same sentence position. This finding cannot be explained by children’s desire to produce tones in a contrastive manner. Instead, it may suggest that children might not compress pitch in the post-focus condition as much as adults do at the age of four. It will be useful to examine this speculation in future research on post focus compression in Mandarin-speaking children’s focus marking.

5. Conclusions

Despite the apparent parallel between music and prosody, the effect of musicality on the development of various levels of prosodic structure has not been studied extensively in the literature. This study investigates the effect of musicality on the development of focus-marking in a tone language, Mandarin. Using tonal dispersion as a proxy for prosodic focus-marking, we compare whether narrowly-focused and non-focused conditions are produced with different degrees of prominence in children between the age of 4 and 6 years. Results show that unlike the robust effect of musicality on the development of tones, the effect of musicality on the focus-marking abilities is quite limited. The only musicality effect found is in the ability to contrast focused and post-focal conditions in the 4-year-olds.

6. References