The Acquisition of Mandarin Tones by Japanese Learners

Hong Zhu¹, Peng Zhang², Kei Yoshimoto¹

¹The Graduate School of International Cultural Studies, Tohoku University, Japan
²Japanese Department, Zhongnan University of Economics and Law, China

zhuhong200812@yahoo.co.jp, yaoyuan2046@znufe.edu.cn, kei@compling.jp

Abstract

This study focuses on the problem Japanese learners have in distinguishing between Tone 2 and Tone 3 when learning tones in Mandarin. Acoustic experiments were carried out based on the four basic vowels in Mandarin. They were designed using continual sound stimuli, which gradually based on the four basic vowels in Mandarin. Acoustic experiments were carried out

This study focuses on the problem Japanese learners have in distinguishing between Tone 2 and Tone 3 when learning Mandarin tones. Acoustic experiments were carried out based on the four basic vowels in Mandarin. They were designed using continual sound stimuli, which gradually

1. Introduction

Although there are plenty of previous studies on the acquisition of Mandarin tones by Japanese learners, these studies almost all focused on comparisons of their two phonetic systems, with particular attention paid to pitch and mora. However, differences in the phonetic systems can not completely account for the difficulties Japanese learners have in learning Mandarin tones - a more in depth study from the aspect of the constituent factors of tones needed to be carried out. Studies on perception of tones in the past demonstrate that F0 contour makes up the basic parameters for distinguishing between tones (e.g., Chuang et al., 1972). Moore and Jongman (1997) also verified the importance of turning point and ΔF0 in their study on the distinction between Tone 2 and Tone 3 by native speakers of Chinese. We thus performed a perception experiment by applying Tone 2 to Tone 3 continua to Japanese learners and Chinese native speakers. The Tone 2 to Tone 3 continua were composed by varying pitch contour tones using Praat across all four of the vowels. The difficulties for Japanese native speakers in learning Mandarin tones were measured by detecting whether the Japanese learners were affected by turning point and/or ΔF0 when learning to discriminate between Tone 2 and Tone 3.

2. Perceptual cues for Tone 2 and Tone 3: Turning point and ΔF0

Earlier studies have considered that the reason tones 2 and 3 may cause the most confusion in perception tests is that they are similar in contour when spoken in isolation. (e.g. Chuang et al., 1972). While overall ΔF0 height may contribute to the distinctive phonetic characteristics of tones 2 and 3, two additional acoustic dimensions are relevant: the timing of the turning point, defined as the duration from the onset of the tone to the point of change in F0 direction, and the decrease in F0 from the onset of the tone to the turning point, hereafter called ΔF0. These properties are schematized in figure 1, below. Perception studies of Mandarin tones 2 and 3 have found that both the timing of the turning point and ΔF0 are perceptually relevant for identification of the tones (Shen and Lin, 1991; Shen et al., 1993). Furthermore, Moore and Jongman (1997) found that there is an interdependency between ΔF0 and the timing of the turning point for Chinese native speakers, which contributes to perception of Mandarin tones 2 and 3. Therefore, this study was based on these two pivots for phonetic synthesis and aimed at determining whether or not the Japanese learners are affected by them when they learn to discriminate between Tone 2 and Tone 3.

Figure 1: Turning point and ΔF0 properties schematized for a contour tone (Moore and Jongman, 1997)

3. Methods

3.1. Subjects

Twenty-seven native speakers of Japanese between 19 and 25 years old who were learning Mandarin and Twenty-seven native speakers of Chinese (20-25 years old) produced the data for this study. None of them have hearing impairments. Subjects were paid for their participation.

3.2. Stimuli

There are essentially 6 single vowels in Chinese, but /o/ and /e/ should be considered diphthongs, and they should be notated as /o/ and /e/ respectively, according to Chao (1968) and Zhu (2010). The present study therefore chose to exclude the controversial /o/ and /e/, and focused only on /a/, /u/, /a/ and /u/ for synthesis.

Vowel pronunciation data taken from a female native speaker (25 years old) from Beijing was used in our experimentation. The recordings were made using a Roland R-09HR recorder. Since the four vowels were played to subjects
randomly, the time of duration and the sound pressure were normalized to ensure the subjects to make proper judgments.

Stimuli were created based on the methods of Moore and Jongman (1997) in their experiment 2, and thus the perception tests were designed to vary the timing of the turning point along a continuum from 20 to 220 ms in 40-ms steps, for a total of 6 stimuli for each vowel. These stimuli were hypothesized to trigger Tone 2 responses when the turning point occurred closer to the tone onset and Tone 3 responses when the turning point occurred later in the tone. In addition, ΔF₀ was varied from 10 to 70 Hz in steps of 15 Hz, generating 5 stimuli for each vowel. Since Tone 2 typically exhibits a shallower ΔF₀, it was expected that tones with lower ΔF₀ would produce more Tone 2 responses than tones with a higher ΔF₀.

Manipulating both the timing of the turning point and the ΔF₀ allowed us to understand how these acoustic parameters are used in Japanese learners’ perception of tones 2 and 3. Table 1 represents all combinations of these parameters that were included in perception experimentation of each vowel. The turning point manipulations are represented along the horizontal axis, and the ΔF₀ manipulations on the vertical axis. According to traditional phonetic descriptions, Tone 2 is characterized by a short fall in F₀ followed by a long rise, while Tone 3 has a deeper, longer fall, followed by a long rise. The dark region corresponds to characterizations of Tone 2, and contains stimuli with turning points from 20 ms to approximately 140 ms, along with lower ΔF₀ values. A high ΔF₀ coupled with an early turning point (20 to 60 ms) could also be expected to yield a perception of Tone 2, since the F₀ rise of the stimulus is predominant. On the other hand, listeners could be expected to perceive any stimulus containing a deeper F₀ fall and a longer duration to turning point as Tone 3, marked in gray in Table 1.

### Table 1. Combinations of turning point and ΔF₀ manipulations for synthesized stimuli.

<table>
<thead>
<tr>
<th>ΔF₀ (Hz)</th>
<th>20</th>
<th>60</th>
<th>100</th>
<th>140</th>
<th>180</th>
<th>220</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There were 240 tokens in total (6 turning point variations × 5 ΔF₀ variations × 4 vowels × 2 repetitions of each). The dark region corresponds to predicted tone 2 responses, and the gray region corresponds to predicted tone 3 responses.

### 3.3. Procedure

Stimuli were played in randomized order using the Praat experiment MFC program. Subjects were asked to determine if they had heard Tone 2 or Tone 3, and click on the according box, shown in figure 2, below. A break was given after every 40 stimulus.

### 4. Results

We used the Residual Analysis of Cross-tabulation test in order to clarify the correlation between subjects’ frequency of selection of tones 2 and 3 about the stimulus, and obtained the frequency distribution shown in Figure 3. The dark region represents a significantly large (p<.05) number of Tone 2 responses. Similarly, the gray region shows a significantly large (p<.05) number of Tone 3 responses. The bold lines show the predicted boundary between Tone 2 and Tone 3. In using a test for independence, we found a significant difference in the native Chinese speaker data when the ΔF₀ and turning point were changed. This shows that there is an interdependency between ΔF₀ and the timing of the turning point in our data, as in Moore and Jongman (1997), which contributes to the perception of Mandarin tones 2 and 3. However, we did not find such an interdependency between ΔF₀ and the timing of the turning point in the Japanese learners.
A discrimination analysis was then performed on the results of the Residual Analysis for Tone 2 and Tone 3 on factors of turning point and ΔF0. The results show the p value of the turning point and ΔF0 in the four vowels is less than 0.05, and that there is thus a difference in the population mean of judgments about tones 2 and 3. We could therefore determine that Japanese learners are affected by both turning point and ΔF0, in their distinguishing between Tone 2 and Tone 3. Furthermore, we found that the absolute value of coefficients in standardized coefficient of Linear Discriminate functions was larger for the turning point, meaning that it is a more important determining factor than ΔF0 for Japanese learners of Mandarin in telling the difference between Tone 2 and Tone 3.

5. Conclusion

The purpose of the perception experiments reported in this study were to determine if and in what ways Japanese learners are affected by the factors of turning point and ΔF0 in distinguishing between Tone 2 and Tone 3 in Mandarin. Our results indicate that they are indeed influenced by both factors, but that turning point is more predominant. Thus, turning point is likely the primary cause of Japanese learners’ confusion in learning Mandarin tones. Furthermore, we found an interdependency between ΔF0 and the timing of the turning point that contributes to perception of Mandarin tones 2 and 3 in our native speaker data, just as was reported in Moore and Jongman (1997).

In addition, Japanese learners’ discriminative distribution diagram, created from our results, illustrates the categorical perception of Tone 2 and Tone 3 is different in four basic Mandarin vowels. According to traditional phonetic descriptions, Tone 2 is characterized by a short fall in F0 followed by a long rise, while Tone 3 has a deeper, longer fall, followed by a long rise. The distribution diagram can be predicted as in Table 1. Compared with Table 1, we found the order of perception ambiguity in the 4 vowels sequence from Japanese learners’ discriminative distribution diagram should be /i/ > /ü / > /a/ > /u/. Meanwhile, the diagram of articulation of Chinese vowels from Shi (2008) also shows the sequences of the F2 values (one of the most important factors to constituent of vowel) of the four vowels to be /i/ > /ü / > /a/ > /u/. Our results thus suggest that the smaller the F2 value, the better the ability of Japanese learners of Mandarin to perceive the difference between Tone 2 and Tone 3. Thus, F2 value may be a significant factor that influences the tone perception of Japanese learners of Mandarin, which warrants future investigation.

6. Acknowledgements

I gratefully acknowledge the help of Professor Yoshio Kitahara. I do appreciate his professional instructions and suggestions on this study. Also, I would like to thank Mrs. Zhang Libo, the Chinese teacher in Tohoku University of Japan, who kindly assisted us during the experiment in this study.

7. References