An Online Perceptual Test of Mandarin Chinese Tones by L2 Learners of Tonal vs. Nontonal Languages

Chin-Chin Tseng 1, Alice Lee 2

1 Graduate Institute of Teaching Chinese as a Second Language, National Taiwan Normal University
2 Aim for the Top University Plan, Ministry of Education, Taiwan
tseng@ntnu.edu.tw, alice66alice66@gmail.com

Abstract

The principal aim of the online Mandarin Chinese perceptual test is to understand how learners of Mandarin Chinese identify the language’s four tones and its correlated factors. There are many factors involved in successfully learning a second language. Therefore, aside from a learner’s L1 language and prior experience, his/her attitude and motivation also contribute to L2 learning. That may be why previous research did not come to a conclusion about how L2 learners acquire Mandarin tones. In this study, we investigated tonal vs. nontonal, and other factors that may contribute to the perceptual test results. An online perceptual test of Mandarin tones is designed for the collection of L2 learners’ perceptual data. We collected data from a total 206 participants in this study. 75 participants identified themselves as tonal speakers, 91 as nontonal speakers, while 40 chose other. The participants were university students in the US and Vietnam, learning Mandarin Chinese at the time they took the experiment. The test contains a questionnaire and 20 test items. The results show that tonal speakers achieved significantly higher accuracy rates in identifying Mandarin tones and their response time was significantly shorter than nontonal speakers. Most identification errors for tonal speakers occurred for Tone 1, while Tone 2 had the most errors for nontonal speakers. Correlation analysis shows that majoring in Chinese is positively correlated with their test score. Regarding attitude and motivation factors, the more important one valued tones, the better he/she performed. The more the student wanted to master Chinese tones, the better he/she performed.

Index Terms: tone, perception, Mandarin Chinese, second language

1. Introduction

Being able to distinguish tones is a key element for learning Mandarin Chinese as a second language. Linguistic transfer has shown that L2 learners tend to generalize in their knowledge about their first language (L1) to help them learn a second language (L2). Moreover, the L1 pitch features such as the stress and intonation systems are generally a major source of difficulty in learning tones, and especially influence American listeners’ perception of Mandarin tones. Michaud (2008) pointed out that the presence or absence of lexical tones in a language does not by itself constitute a fundamental typological divide for intonation systems. Tonal languages use pitch to distinguish lexical meaning, while nontonal languages use pitch to express emotional and other paralinguistic information. In East Asia, Chinese and Vietnamese are tonal languages. English is considered a typical nontonal language. Therefore, we invited college students from the US and Vietnam to participate in the online perceptual test. Our assumption was that Vietnamese students already possessed tonal categories, and that they should require less time to identify a tonal category. Moreover, we assumed they would perform better than nontonal speakers at the beginning level since nontonal speakers are less sensitive to lexical tonal categories, and they have to build their categorical perception for lexical tones from scratch.

2. Literature Review

For speakers whose native language is nontonal, tone has presented great difficulty since the functional association between the F0 characteristics and the segmental structure is unfamiliar to them (e.g., Kiriloff, 1969; Shen, 1989). Lee et al. (1996) found that Cantonese speakers (tonal language) did better than English speakers (nontonal language), at discriminating Mandarin tones. Wang, Jongman, and Sereno (2001) showed that Mandarin tones are predominantly processed in the left hemisphere by native Mandarin speakers, while native speakers of American English with no prior experience with a tone language did not show dominance in either hemisphere.

Wang (2006) summarized the perception of Mandarin tones investigated in terms of the identification of the tonal categories. It was found that non-native listeners’ tone perception tended to be less categorical as compared to that of native listeners. Listeners from tonal languages can perceive tonal contrast better than those from nontonal ones. The source of difficulty in tone acquisition has often been attributed to non-native speakers’ lack of sensitivity to tonal categories. A second factor is the non-native speakers’ lack of experience in processing the various phonetic features that characterize Mandarin tones. Non-native speakers attached more importance to pitch level, instead of the pitch contour dimension of F0 as native speakers did. These results about the processing of Mandarin tones for native speakers raise the question as to whether non-native speakers process the tones auditorily (i.e., based on innate psychoacoustic mechanisms) or linguistically (i.e., resulting from language-specific experience). In our study, we minimize the interference of segmental effect on tones, and reduce the perceptual item to a single syllable to minimize the interaction of stress and intonation on tones. The results show that speakers whose native language is nontonal, tone still presented certain difficulty, since the functional association between the F0 characteristics and the tonal category is unfamiliar to them.

So (2006) examined to what extent phonological and phonetic influences from native prosodic systems affected the perception of non-native lexical tones. The results showed that native tonal experience may hinder its speakers’ ability to discriminate tonal differences at the phonetic level. In our
study, the perceptual results of Vietnamese speakers coincide with their pronunciation of Tone 1 (a high level tone). In the misjudgment of Tone 1, 75% were identified as Tone 4 (a falling tone), because their pitch level for Tone 1 is slightly lower than Mandarin Chinese, and their pitch tends to fall at the end.

3. Method
An online perceptual test of Mandarin tones was designed for the collection of L2 learners’ data. The test contains a questionnaire and 20 test items (Fig 1 & 2).

3.1. Participants
206 university students in the US and Vietnam, learning Mandarin Chinese at the time participated in the experiment. 75 participants identified themselves as tonal speakers. 91 identified themselves as nontonal speakers, and 40 chose other.

3.2. Stimuli
In order to minimize the effect of segments on tone perception, the syllable [a] was used. Lin (2006) and Wang (2012) both found that L2 learners of Mandarin Chinese have greater difficulties in identifying the correct tone category when there exists segments which do not exist in their first language. The stimuli were spoken by a female native speaker of Mandarin from Taiwan. Table 1 is the F0 measurement of the perceptual stimuli.

Table 1. The F0 of the perceptual stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>333-356</td>
</tr>
<tr>
<td>T2</td>
<td>221-371</td>
</tr>
<tr>
<td>T3</td>
<td>194-79-265</td>
</tr>
<tr>
<td>T4</td>
<td>387-131</td>
</tr>
</tbody>
</table>

Figure 4: Graph of T1 to T4 converting F0 into Log representation

Each tone was recorded in a professional studio. Figure 4 shows the representation of the four stimuli (T1 to T4). The stimuli were checked by four native speakers of Mandarin for the appropriateness of representing the four tones in Mandarin Chinese. Each tone was copied five times, and randomly arranged into the 20 test items.

3.3. Procedure
Each subject was instructed to complete an online questionnaire regarding their experience, attitude, and motivation before the tone perception test. This tonal test has 20 test items. Each item is played twice. Subjects were told that response time would also affect their final score, meaning the less time spent answering the questions correctly would add points to their final score. After hearing the [a] sound, they were asked to identify the correct tone. After completing an item, the subject clicked on the “next” button to continue to the next item. The test score and the response time appeared on the screen when the test was completed. See Figure 3.

3.4. Data Analysis
Based on the number of correct responses, the mean scores of 75 tonal speakers and 91 nontonal speakers were calculated and compared using independent sample t-test. In addition,
we calculated the percentage of accuracy for each test item based on the performance of each group. One-Way-ANOVA was then used to test the difference between each tone.

4. Results

The results show that the mean score for tonal speakers is 19.91 and 18.18 for nontonal speakers. Tonal speakers achieved significantly higher accuracy rates ($t=5.533$, df=94.174, $P<0.01$) in identifying Mandarin tones.

**Figure 5:** Mean Score of accuracy between tonal (T), nontonal (NT) and other Group (O)

The mean response time was 41.72 seconds for the tonal language group, and 49.84 for the nontonal group. The difference in response time between the two groups reached significant level ($t=-4.251$, df=152.853, $P<0.01$).

**Figure 6:** Mean Score of response time between the tonal (T), nontonal (NT) and other groups (O)

The most frequent identification error for nontonal speakers occurred with Tone 2. The collected data was analyzed with One-Way-ANOVA. The test score of Tone 2 is significantly different from other tones among nontonal speakers. $F(3,16) = 3.271$, $P<0.05$.

**Figure 7:** Accuracy rate of four tones (NT speakers)

Tables 2 to 5 are error analyses for each tone misjudged by nontonal speakers. There were 46 tokens of Tone 1 mistaken for Tone 2, which was 90.2% of all Tone 1 errors. In Figure 4, we can see how stimuli Tone 1 and Tone 2 look similar. The major visual difference is the pitch level.

**Table 2. Tone 1 Error Analysis (NT speakers)**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>46</td>
<td>90.2</td>
</tr>
<tr>
<td>Valid</td>
<td>4</td>
<td>9.8</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100.0</td>
</tr>
</tbody>
</table>

There were 58 tokens of Tone 2 mistaken for Tone3, which made up 73.4% of all Tone 2 errors. Tone 2 and Tone 3 do not look alike in Figure 4; however, they seem to share similar beginning and ending pitches compared with the other tones.

**Table 3. Tone 2 Error Analysis (NT speakers)**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>22.8</td>
</tr>
<tr>
<td>3</td>
<td>58</td>
<td>73.4</td>
</tr>
<tr>
<td>Valid</td>
<td>4</td>
<td>3.8</td>
</tr>
<tr>
<td>Total</td>
<td>79</td>
<td>100.0</td>
</tr>
</tbody>
</table>

There were 15 tokens of Tone3 misjudged as Tone2, which made up 68.2% of all Tone 3 errors. (Table 3)
There were 11 tokens of Tone 4 mistaken for Tone 1, which made up of 78.6% of all Tone 4 errors. In Figure 4, we see Tone 4 and Tone 1 share a similar starting pitch.

For tonal speakers, the results of One-Way-VOVA show no significance between the four tones among tonal speakers. This might be because the accuracy for all four tones was close to 100%. However, the distribution of error patterns is different for tonal and nontonal speakers. Tonal speakers seem to have a much better command when it comes to identifying Tone 2.

Correlation analyses show that the shorter the response time is, the better perceptual score one gets; moreover, the stronger the emotional reaction is when a person first encounters Mandarin tones, the higher perceptual accuracy one gets.

For subjects not fond of music, their response time is longer, and their performance is poorer than those who like music.

Regarding the attitude and motivation factors, the more a person values tones, the better they scored overall (r=.195, p<0.05). The more one wants to master Chinese tones, the better they performed (r=.265, p<0.05).

### 5. Discussion

So (2006) conducted a tone discrimination experiment on Cantonese (a tonal language) and Japanese speakers (a nontonal language). In So’s study, the results show that listeners’ sensitivity to Tone 2 was significantly lower than their sensitivity was for the other three tones. In our study, nontonal speakers also scored low on Tone 2 identification; however, our tonal speakers (Vietnamese speakers) achieved 100% accuracy on Tone 2 identification. Vietnamese students only have a few errors with Tone 1 (high even level tone). 75% of the Tone 1 errors were mistaken for Tone 4.

So (2006) reported that Cantonese listeners made significantly more errors for the T1-T4 pair. Cantonese and Vietnamese are both tonal languages. Tseng (2008) reported an analysis on the perception of Mandarin Chinese tones by L2 learners from Southeast Asia, the results of which showed that out of the 600 test tokens for each tone, the error rate of identifying T1(24%) and T2(23%) are higher than T3(13%) and T4(16%). Vietnamese speakers and Cantonese heritage speakers in Malaysia also misidentify T1 for T4.

### Table 4. Tone 3 Error Analysis (NT speakers)

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Valid</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>22.7</td>
<td>22.7</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>78.6</td>
<td>78.6</td>
</tr>
</tbody>
</table>

### Table 5. Tone 4 Error Analysis (NT speakers)

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Valid</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>78.6</td>
<td>78.6</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8: Accuracy rate for the four tones (tonal speakers)

Correlation analyses show that the shorter the response time is, the better perceptual score one gets; moreover, the stronger the emotional reaction is when a person first encounters Mandarin tones, the higher perceptual accuracy one gets.

For subjects not fond of music, their response time is longer, and their performance is poorer than those who like music.

Regarding the attitude and motivation factors, the more a person values tones, the better they scored overall (r=.195, p<0.05). The more one wants to master Chinese tones, the better they performed (r=.265, p<0.05).

Although our study seems to suggest that linguistic experience of tones does facilitate the perception of non-native tones, we refrain from making such conclusions because the
results might be contributed to the transfer of training. Subjects in Vietnam National University, Hanoi have on the average 10 more hours of Chinese lessons per week than students in Penn State University in the US. Furthermore, at VNU the introduction to Mandarin pronunciation (including tones) is composed of four lessons, while at Penn State an intensive pronunciation lesson is offered at the very beginning, and the identification of correct tones is mainly practiced in the first year of learning Mandarin Chinese, instead.

Native languages definitely play an important role in the perception of non-native tonal contrasts, because the same stimuli were perceived differently by learners of different L1. Moreover, several studies show that Vietnamese speakers have greater difficulties in identifying Mandarin T1.

Wang et al. (1999) raises the question as to whether the processing of Mandarin tones for advanced learners can be authentically native-like as the learners achieve high proficiency in Mandarin. In our study, the accuracy rate seems to slide downward with the year of learning.

Wang (2006) mentioned that the source of difficulty in tone acquisition has often been attributed to non-native speakers’ lack of sensitivity to tonal categories. We found that tonal

![Figure 11](image1.png)
**Figure 11:** Mean Score of first year students between American (A) vs. Vietnamese (V) Group

![Figure 12](image2.png)
**Figure 12:** Mean Score of second year students between American (A) vs. Vietnamese (V) Group

![Figure 13](image3.png)
**Figure 13:** Mean Score of third year students between American (A) vs. Vietnamese (V) Group

Previous research indicates that the brain’s left hemisphere is more adept at phonemic processing, including phonemes, syllables, and words (Kimura, 1961; Studdert-Kennedy and Shankweiler, 1970), while the right hemisphere is better at melodic and prosodic processing, including music, pitch contours, and affective prosody (Kimura, 1964; Curty, 1967; Bryden, 1963). The function of Mandarin tones is similar to phonemes, but the realization of tones is closely related to pitch contour. In our study, participants were asked to choose from “never”, “under 3 month”, “3 months to a year”, “1-3 years”, and “more than 3 years,” to indicate their experience in learning musical instrument or sing. A score (from 1 to 5) were assigned to each time course. The mean score for each group were calculated and compared using independent t-test. The result showed that 61.4% of the tonal speakers never studied any musical instruments or took vocal lessons, while 57.1% of the nontonal speakers had more than three years of musical training. There is a significant difference between the two groups (average 1.68 points for tonal speakers and 3.99 points for nontonal speakers, t=2.205 , df=66.828, P<0.05).

We are not sure if this factor has affected the performance of the nontonal speakers. Why is it suggested in various studies that it is harder to correctly identify Tone 2 than any other tone? This can be an important direction for future research.

Further study needs to be done on subjects belonging to the other category. We often ask the subject to identify their native language for the purpose of second language research, but in reality more and more people are bilingual or multilingual, or travel and reside in various language environments, making it difficult to isolate L1 factors that may influence their L2 learning. As we know, a significant percentage of students taking Mandarin lessons in the US have a Chinese ancestral background. How should they be categorized? In a strictly experimental design of studies, these subjects will be excluded because they do not conform to the required conditions. However, being a language educator, we emphasize learner-centered education. We need to understand how each learner is processing their L2 system. We need more individual longitudinal studies to have a clear picture of how tones are perceived, because after all, perception exists in every individual.

### 6. Conclusions

Wang (2006) mentioned that the source of difficulty in tone acquisition has often been attributed to non-native speakers’ lack of sensitivity to tonal categories. We found that tonal
language speakers were better at correctly identifying tones in terms of speed and accuracy of their responses than nontonal language speakers. Tonal language speakers seemed to acquire general tone identification abilities; thus, it seems that a listener’s strategy for tone perception depends to a certain extent on the linguistic function of pitch in their native language.

Although nontonal and tonal speakers process Mandarin tones differently, our study confirms that non-native learners’ ability to identify the tones can be significantly improved after a short period of learning. Wang, et al. (1999) pointed out that adult L2 learners can establish separate phonetic categories for those L2 sounds that are nonexistent in their L1 sound systems. Lin (2006) reported a longitudinal study on the acquisition of Mandarin tones by rudimentary Japanese learners. Her results show that beginning Japanese learners can form categorical perception of tones within three months of learning Mandarin Chinese, though the result is still not perfect. In our study, nontonal speakers also demonstrated their ability to achieve fairly high scores after they took Mandarin classes for three months. However, as the years of learning progress, their scores show a slightly downward trend. This raises the question as to whether the processing of Mandarin tones for advanced learners can be authentically native-like as the learners achieve high proficiency in Mandarin. Therefore, one important direction for future research is to systematically investigate the learning of Mandarin tones as a function of proficiency for a better understanding of the mechanisms underlying the dynamic process of language learning at the suprasegmental level.

7. Acknowledgements

This research project was sponsored by the Ministry of Education in Taiwan (NTNU’s Aim for the Top University Project). Many thanks to students, teachers and the Chinese coordinator at Penn State University; students, teachers and the Department of Chinese Language and Culture at Vietnam National University. This research could not have been accomplished without our research team members. This research required university and industry liaison systems and in-depth interdisciplinary research, including speech perception, second language acquisition, and an automatic online assessment platform. Special thanks to Lu Feng Technology Inc. for the test platform support. We would like to thank the organizing committee of TAL 2012 for supporting this special session on acoustic, articulatory and perceptual study and new technology for L2 learning of Mandarin tones.

8. References