Identification of Mandarin Tone-2 and Tone-3 in Disyllabic Contexts by Chinese Natives and Japanese Students

Wen Cao, Yuko Ota, Hong Jiang, Lu Liu and Jinsong Zhang

Center for Studies of Teaching Chinese as a Foreign Language, Beijing Language and Culture University
tsao@blcu.edu.cn, yuko.otankonasu@gmail.com, rainbowjiang19@hotmail.com, liulu1228@gmail.com, jinsong.zhang@blcu.edu.cn

Abstract

This study primarily explores the perception of Mandarin Tone-2 (T2) and Tone-3 (T3) in disyllabic contexts by Chinese natives and Japanese students. Seventeen contours of a T2-T3 continuum and twelve contours of a T2-‘Half-T3’ continuum are respectively synthesized and connected before or after the four tones separately as stimuli. The results of their perception show that: (1) Both Chinese and Japanese participants are affected by contexts in tone identifying; (2) Japanese rely more on acoustic characteristics other than phonological rules; (3) T3 identification is more difficult than T2 for Japanese, especially when it is after the context tone.

Index Terms: tone, perception, context, Chinese, Japanese

1. Introduction

Although it has been reported that phonation types have an effect on the perception of tones [1, 2], the acoustic performance of the tones of the Standard Chinese or Mandarin is believed mainly to follow the pattern of f0, which is also the major clue in its tonal perception [3, 4]. Besides, in practical running speech, the f0 pattern has more or less deviation due to coarticulation. For example, there are two tone sandhi rules for T3 in Mandarin: ‘T3→T2 /__+T3’ and ‘T3→Half-T3 or Low-falling /__+{T1, T2, T4}’. Native speakers neglect those deviations and regard them as the same phoneme. However, tone context does have an effect on Mandarin tonal identification. As is mentioned in [3], native speakers have resort to two clues in tonal perception — pitch variation itself and that of the adjacent tones, i.e., tone contexts.

Japanese who learn Chinese as a second language (L2) have been found having difficulties in tone perception, especially in distinguishing T2 (LH, the rising tone) and T3 (LL, the concave or dipping tone) [5, 6]. Lee et al found that in tonal perception, native Chinese rely more on tone contexts while English speaking learners on the tone itself [7, 8].

What about Japanese (L1) learners of Chinese (L2)? Do they rely on tone contexts or the tone itself? And if yes, how do they perceive Chinese tones under different tonal circumstances? This research is trying to answer these addressed questions and to explore more knowledge on Chinese tones from the aspect of Chinese natives’ perception.

2. Method

2.1 Stimuli and subjects

The original materials are taken from 15 samples of Mandarin disyllabic tonal combinations with nonsense meaning — “dada” by a female Chinese native speaker. The f0 contour based on the average pitch value of the post “da” in T2 is taken as the target T2, which is obviously a rising tone; and that based on the average pitch value of the post “da” in T3 is taken as the target T3, which is a dipping tone. However, the average pitch value of the preceding T3, which is actually realized as a low-falling tone, is taken as the target of the sandhi ‘Half T3’ (hereinafter to be mentioned as T3*). Each syllable is set into the length of 350 ms. The two continua, T2-T3 and T2-T3* are made through 17 steps and 12 steps separately, modified accordingly in a 95% confidence interval for the values of pitch at duration points of 25% (five points, namely 0%, 25%, 50%, 75% and 100%).

Figure 1: Tone contours for T2-T3 continuum (left) and those for T2-T3* continuum (right).

Figure 2: T2-T3 continuum under eight tone contexts.
Eight kinds of disyllabic tonal combination are obtained (X+T1, X+T2, X+T3, X+T4, T1+X, T2+X, T3+X, T4+X), with X presenting one of the contours of the continua. There are totally 232 disyllabic samples for stimuli.

Aged 18 to 35, thirteen students from eastern Japan, who have learned Chinese for more than 2 years, were enrolled to take part in the experiment. Ten Chinese students, aged 20 to 30, all native Beijing speakers of Mandarin, served as the control group.

2.2 Procedures

Sounds were played by a computer automatically while participants filled out a questionnaire to describe the tonal types they heard. Each disyllabic sample just played in random arrangement once, with a 4.5 second interval in between.

All listeners had a short period of orientation before test, in which they were given 20 samples.

3. Results and discussion

3.1 On the T2-T3 continuum

![Figure 3: The identification result of T2-T3 from Chinese.](image)

Figure 3 displays the result of tonal identification from Chinese natives, where the perceptual boundary between T2 and T3, which seems to exist in Figure 3, seems to be changing — from the sixth contour to the ninth. Repeated Measure ANOVA shows that there is a significant tendency of the interaction between tone contexts and the identification of Tone-2 and Tone-3 (F_{T2} (1, 7) = 6.44, p<0.05; F_{T3} (1, 7) = 6.10, p<0.05). It proves that the perceptual boundary depends on the tone contexts. To make it more distinct, the perceptual boundary between T2 and T3 emerges later along the continuum when the test syllable possesses the post position other than the preposition (F (1, 3) = 21.80, p<0.05). This means that the pitch contour following a tone needs to be more concave if it is to be perceived as a T3.

Precisely, the perceptual boundary between T2 and T3 varies accordingly:

- It is at the 9th contour in T1+X;
- It is between the 7th and the 8th in T2+X;
- It is between the 6th and the 7th in T3+X, T4+X and X+T1;
- It is between the 5th and the 6th in X+T2;
- It is at the 6th contour in X+T3 and X+T4.

In spite of the difference mentioned in the former paragraph, we discover some similarity: the perceptual boundary between T2 and T3 is later/lower when the preceding context tone ends with [H] other than that with [L] (T1+X and T2+X vs. T3+X and T4+X), and it is later/lower as well when the following context tone starts with [-L] other than that with [L] (X+T1, X+T3 and X+T4 vs. X+T2).

So, it can be induced that no matter a context tone is before or after a test pitch contour, an [L] tone featured point right next to it will preset the perceptual boundary between T2 and T3.

What should be noted is that most of the contours including and after the 7th of the T2-T3 continuum, which are concave and dipping, are identified as T3(*) by most of the Chinese participants even when they appear in the front of the disyllable (i.e. X+T1, X+T2, etc.). It indicates that the Tone-3 sandhi rules are not obligatory or must-be rules, but conventional or used-to-be. Actually, in simultaneous running speech of Chinese, concave or dipping Tone-3 tokens are often observed.

Relationally, what is also worthy to be mentioned of in the test is that most Chinese participants (8 out of 10) perceive the ‘T3+*X’ pitch contours after the 8th as T4+T3. It may be caused by two reasons: (1) Conflicts between the conventional sandhi rule ‘T3→T2 /_+T3’ and the falling pitch contour preceding; (2) Both T4 and T3* have falling pitch contours.

It can be seen from Figure 4 that the perceptual boundary between T2 and T3 from Japanese students is not as evident as Chinese participants. However, significant interaction is also found between tone context and perceptual identification (F_{T2} (1, 7) = 9.01, p<0.05; F_{T3} (1, 7) = 14.04, p<0.05). Repeated Measure ANOVA proves that, the contours of the T2-T3 continuum in T4+X combination are more likely to be identified by Japanese students as T2 compared to other combinations. Similar results can be seen in the part of T4+X.

It seems extremely difficult for Japanese participants to recognize T3 when it is after a H tone (T1) or HL tone (T4), no matter how dipping or concave it is. The situation seems to be better if it is after a T2 or T3*, easy to compare maybe. However, there is surely a bundle of confusing contours in T3+X circumstance (from the 8th to 12th) for Japanese students to judge tones by.

On the contrary, when the contours of the T2-T3 continuum are before the context tone, they have better chance to be identified by Japanese students as T3. When we compare the results of T1+X and T4+X with X+T1 and X+T4, situation will be relatively clear. The identification result that most similar to Chinese participants’ is that of X+T2, with boundary between T2 and T3, not as certain as Chinese though. What can be induced here is that it will be less
difficult for Japanese students to learn tones before a tone context than after it.

It is obvious that, the tone perception is vulnerable to the tone contexts for Japanese students, too; different from that for Chinese though.

![Figure 4: The identification result of T2-T3 from Japanese.](image)

Figures 5 and 6 depict the result profiles of T2 and T3 identification by Chinese and Japanese participants under different tone contexts. It is clear that the two groups of the participants have different performance in T2 and T3 identification. The consistency in tonal-context-dependent T2-identification is higher among Chinese students than in Japanese group. So is the T3-identification. For Japanese students, the pitch range for perceiving a tone as T2 is broader than that for T3. Although it seems that T2-identification is easier for Japanese students, it doesn’t mean they have built up identical perceptual space of T2 as the native speakers.

![Figure 5: Tendency (profile) for T2-identification in T2-T3 continuum by Chinese (left) and Japanese (right).](image)

3.2 On the T2-T3* continuum

The result of the identification on the T2-T3* continuum by Chinese is obviously more complicated than that on T2-T3, which can be seen from Figure 7. The contours are recognized as from T1 to T4 separately in various tone contexts, different from the identification result of T2-T3 (cf. Figure 3). Maybe it is because the T2-T3* continuum is composed of contours of from rising to falling, instead of from rising to dipping.

![Figure 7: The identification result of T2-T3* from Chinese.](image)

The statistic demonstrates that there is a significant interaction between tone-identification and tone contexts: $F_{T1}(1,7)=12.37, p<0.05$; $F_{T2}(1,7)=13.51, p<0.05$; $F_{T3}(1,7)=48.02, p<0.05$; $F_{T4}(1,7)=17.34, p<0.05$.

There is little possibility for Chinese to identify the test syllables as T1 in X+T1, X+T2 and X+T4. However, contours from the fourth to the seventh of the T2-T3* continuum in the
rest tone contexts are perceived as T1 by most Chinese participants. Through the Repeated Measure ANOVA, a gradient from easy-to-difficult pattern is found: T3+X*X+X*T3 >T2+X*X+T1+X*T4+X.

In fact, each of the contours among from the fourth to the seventh of the T2-T3* continuum has a pitch change between 2 semitones (cf. Figure1), rising or falling. On one hand, we found that a pitch contour with small slope is still easy to be heard as a T1, which is phonologically described as a level tone; on the other hand, tone contexts do make some difference.

Figure 7 also shows that most Chinese participants identify the first three contours along the T2-T3* continuum as T2 in all eight tone contexts, whereas they express inconsistency after the third contour. In the cases of the situation that X is perceived as T2, the identification difference between T1+X and T3*+X is significant (p<0.05), so is that between T1+X and X+T3. There is no significant difference in other tonal circumstances.

The first three pitch contours of the T2-T3* continuum rise separately with equal-to or more-than 3 semitones. This should be the main reason Chinese recognize them as T2. The significant identification difference between T1+X and T3*+X is mainly caused by their perception on the fourth contour, which is regarded by most Chinese as T2 in T1+X but not as in T3*+X. So is that between T1+X and X+T3. It seems that low tones (T3 and/or T3*) as contexts decrease the rising of the fourth contour in perception.

As for the situation that X is perceived as T3(*), there is a significant difference in the identification of T3(*) between in the tone contexts of X+T1, X+T2, X+T4 and other circumstances (p<0.05). Chinese participants start to perceive the test contours in these three contexts (+T1, +T2, +T4) as T3 (*) from the fifth contour along the synthesized T2-T3* pitch continuum. All contours after the fifth are falling. And actually, it is almost the perceptual boundary contour between T2 and T3 (*) in the mentioned three tonal contexts.

In the combination of T1+X, T3+X, T4+X and X+T3, the test contours from the eighth or the ninth to the twelfth are (1, 7) = 2.29, p>0.05). The only exception is on T3 (*) (F4(1, 7) = 8.64, p<0.05). As for the situation that X is perceived as T3(*) by the Japanese participants, there is a significant difference in the identification between in the tone contexts of X+T1, X+T2, X+T4 and other circumstances (p<0.05). This result is quite similar to that of Chinese. It confirms the point that to learn Chinese tones before a tone context is better than after it for Japanese students.

Comparing with the identification result of T2-T3* from Japanese with that from Chinese, there are still some other information in Figure 8 worth mentioning.

As for instance, Japanese students tend to identify contours from the fourth to seventh of the T2-T3* pitch continuum as T1 in the combination of T3*+X and X+T3. This result is partially similar to that of Chinese. Even that the fifth and the sixth contours in T4+X have more opportunity to be recognized as T1 is also shared by Chinese and Japanese participants. However, some contours out of those four can be perceived by Chinese as T1 in T1+X and T2+X (cf. Figure 7) whereas not by Japanese. This suggests that the H tone feature has stronger effects on Japanese students’ perception of T1. A rational interference is that it will not be accepted by Japanese participants as a T1 (i.e. HH) if a pitch contour is not as high as the contour shape when T3* appears in the unconventional environments.
enough as the foregoing contextual H. Similar finding was reported before in [9].

As to the T2-identification, what needs to be pointed out is that in some circumstances, such as X+T1 and X+T4, Japanese participants identify rising contours as T2 less than Chinese. The reason could be that the following H tone feature raised their expectation on the necessary rising threshold of T2.

Besides, there is significant difference between the identification results of Chinese and Japanese on T1+X. Their recognition on the contours of from the 8th to the 12th in the T2-T3*continuum need to be noted. Most of Chinese participants take them as T4, while the majority of Japanese take as T3 (*). It indicates again that Chinese identify tones on contour shape or pitch direction, but Japanese on register or pitch height.

Figure 9 exhibits the T3 (*) identification under varied environments of both Chinese and Japanese students. It implies that Chinese natives are influenced in identifying T3* by the conventional phonological rules. Chinese natives show a more significant diversion between the identification results from X+T1, X+T2, X+T4 and those from other tonal environments, where T3* will never appear canonically. But Japanese students, in reverse, perceive test syllables as T2 or T3 even in T1+X. This renders an inference that Japanese students rely more on acoustic characteristics (specifically the average pitch height) than phonological rules when they make a judgment while Chinese students have a strong tendency to perceptually categorize the sounds. They have not only acoustic features but also the phonological rules in mind.

4. Conclusion

The T3 sandhi rules are conventional. On one hand, a prototype of the dipping tone without becoming rising or low-falling according to the rules can still be acceptable and recognized as a T3. On the other hand, the chance that a low-falling tone can be taken as the T3 happens almost only in the situation of being preceding. Chinese natives can adopt the rules in depending while identifying tones, but Japanese can not adopt them in some cases.

However, the results of the experiment have shown that both Chinese natives and Japanese learners are affected by tone contexts in perceiving Mandarin tones. Different from English people, as is studied in [8], Japanese learners also rely on tonal contexts while identifying tones, although they may have wrong identifications. The conventional tone sandhi rules and pitch direction impose a stronger constraint on Chinese participants while Japanese students rely more on acoustic characteristics, especially the pitch height, other than being restricted to phonological and systematic tonal rules in perception. T3 identification is relatively the most difficult for Japanese students to learn. However, if it is followed by another tone ([-T3]), it would be easier to be recognized by Japanese learners.

5. Acknowledgements

This study is supported by NSF (Grant # 61175019/F030404; “Research on the model of speech recognition based on the perception of sound variation”).

6. References