A Comparison Study on F0 Distribution of Tone 2 and Tone 3 in Mandarin Disyllables by Native Speakers and Japanese Learners

Ting Zou [1], *Jinsong Zhang* [1,2] , *Wen Cao* [1]

[1] Center for Studies of Chinese as a Second Language, Beijing Language and Culture University
[2] College of Information Science, Beijing Language and Culture University

`neverland-1017@hotmail.com, jinsong.zhang@blcu.edu.cn`

Abstract

In the process of Mandarin learning, Japanese students always have problems when distinguishing Tone 2 and Tone 3, especially in connected speech. So, in this paper, we investigate the production of these two tones in disyllable words on the basis of tone-dependent F0 distribution, trying to show the characteristics of native speakers and the systematic error pattern of Japanese. Getting the error pattern of a specific learner group is of great importance in improving teaching method of Mandarin disyllables, and developing computer assisted mandarin pronunciation technology.

Index Terms: Mandarin disyllables, tone production, Japanese learners, F0 distribution, tonal range

1. Introduction

Fundamental frequency (F0) is often taken as the main acoustic feature of tone identity, although other acoustic properties can also function as perceptual cues, providing "important secondary auditory cues for the identification of discrete tones [1]." As a tone language, the contrast in pitch height and direction in Mandarin can be used to minimally distinguish the meaning of words.

The information carried by tone is realized through regular intra-speaker acoustic change. There may be some inter-speaker variances among native speakers, but the internal acoustic systems of different speakers have a consistent pattern. While for L2 learners, their productive pattern may differentiate significantly from native speakers, showing a consistent error.

In the process of Mandarin learning, lexical tone has presented great difficulty to Japanese learners, for they have to learn to make use of pitch contrasts at the lexemic level. Findings related to lexical tone produced by Japanese learners showed that among the Mandarin tone inventory, Tone 2 and Tone 3 are most problematic[2, 3]. Therefore, it is necessary to investigate the characteristics of these tones in disyllables produced by native speakers and Japanese learners.

1.1. Production of Tone 2 & 3 by native speakers

Previous studies have almost formed a consensus for the F0 contours of Mandarin tones in isolated syllable. Tone 1 is high-level, Tone 4 high-falling. Tone 2 is always regarded as a rising tone, while acoustic analyses often show that there is a slight initial-dipping part, which will not cause tone misidentification. In isolated syllable, Tone 3 is a low-falling-rising tone [6] with a similar contour of Tone 2 but lower and having a later turning point.

In connected speech, Tone 3 has different variants. Chao [4，5] found that Tone 3 always appeared as "half third tone" in running speech, with only a low-falling contour shape. Other studies also showed that in non-prepausal position, Tone 3 remained its "low" feature, but become "half third tone" [6]. The similar features sharing by these two tones and rich variants of Tone 3 also cause some difficulties for native

speakers. A study on tone acquisition of Chinese children found that Tone 2 and Tone 3 were substituted for each other throughout the early learning process [7].

1.2. Production of Tone 2 & 3 by Japanese learners

Some studies [2, 3] have indicated that Japanese students tend to show regular pattern of errors in Tone 2 and Tone 3—in isolated syllable, the rising part of Tone 2 is not high enough and the dipping point in T3 is not apparent.

In disyllables, Tian's experiment showed that "Tone 3+Tone 2" was the most difficult combination for Japanese learners [3]. The "half third tone" was hard to acquire and was always substituted by a level or rising tone in the first syllable. Yuko [8] found that in disyllables, the error rates of Tone 2 and Tone 3 are much higher than Tone 1 and Tone 4, with most errors coming from Tone 3 in the second syllable. In most cases, Tone 3 was produced as Tone 2.

The high inaccurate rates of Tone 2 and Tone 3 by Japanese learners may indicate that generally Tone 2 and Tone 3 are more difficult than the other two tones, and evidence also comes from some studies about acquisition order of Mandarin tones. Table 1 is the relative difficulty of production of four Mandarin tones by L2 learners (adapted from [9]).

Table 1. *Relative difficulty of production of four Mandarin tones reported in previous studies.*

Study	Learners	Order
Leather (1990)	Dutch	T1<T4<T2=T3
Elliot (1991)	American	T1<T4<T2<T3
Chen (1997)	American	T1<T4<T2<T3
Miracle (1999)	American	T1<T4<T3<T2

Zhang [10] carried out a one-year follow-up investigation of Japanese students, recording data about the tone acquisition process. For the beginners, Tone 2 always had a mid-level contour shape and could not be distinguished from Tone 3. After a period of training, there was notable improvement in Tone 1, Tone 2 and Tone 4, while for Tone 3, the fossilization phenomenon appeared.

To sum up, previous studies about Japanese learners' tone production can be divided into two aspects: general description based on teaching experience and empirical studies with acoustic analysis, with a conclusion that confusion of Tone 3 and Tone 2 (substituting Tone 3 with Tone 2) is a common error for Japanese learners and fossilization phenomenon of Tone 3 is notable.

However, for the need of developing computer assisted pronunciation training technology, we still need a more quantitative image about the relative position of each tone within an individual speaker's F0 range, the consistency among native speakers when realizing and maintaining the distinction of the tones as well as the detailed error pattern of Tone 2 and Tone 3 among Japanese learners.

Therefore, this paper investigates the characteristics of tone production by Chinese speakers and Japanese learners.

the basis of intra-speaker F0 distribution. The inter-speaker differences of F0 distribution among native Chinese speakers can give a full picture of the realization of Mandarin tones; the comparison between F0 distribution of Chinese and Japanese learners will show the learners' errors (especially in the error-prone Tone 2 and Tone 3), while the F0 distribution of Japanese learners with different Mandarin experience can shed some light on the development of tone acquisition.

2. Methods

2.1. Material

In disyllable, there are totally 16 tonal combinations. We chose 5 disyllable words for each combination and got a word list consisting of 80 disyllable words.

2.2. Subjects

Twelve Japanese students (6 males; 6 females) with 1-4 years of Mandarin learning experience and 5 Chinese native speakers (5 females) participated in this experiment. The subjects were students in Beijing Language and Culture University, and the two groups were similar in age. All the native speakers were from Northern China speaking standard Mandarin.

2.3. Procedure

The recording was conducted in the phonetics lab of BLCU using Adobe Audition 2.0.

The tone accuracy of the Japanese learners was later checked by 3 native Chinese (students of phonetics in BLCU). In order to avoid the interference of lexical meaning on tone identification, we filtered the segmental information using low-pass filtering method (0-800Hz). Then the 3 native speakers were asked to label the tones of each disyllable combination with a forced choice among four lexical tones in Mandarin. If all of the three listeners disagreed with each other, then that label would not be used; if two of them had identical choice, then that choice would be adopted as the label of the syllable.

Previous research has demonstrated that there are articulatory transition courses at onset and offset of tone contour which are influenced by tonal context and would not affect listeners' perception [11, 12]. So, to present the tone contours in disyllables, we firstly labeled the steady part of tone contour for each syllable.

On the basis of F0 data of steady part in disyllables, two kinds of figures were produced: the figures of overall disyllable pitch contours of two groups and the F0 distribution histograms aiming at showing the full picture of the realization of each tone within individual speaker's F0 range. For the former, we abstracted the F0 value of five points (0, 25%, 50%, 75%, and 100% of the duration) in each syllable and transferred the F0 value to z-score for inter-speaker normalization. Then the average z-score values for these two groups were gained. For the latter, the F0 values of each syllable was extracted every 10ms using PRAAT 5.1.43 [13], and then the tone of each syllable was labeled. For each speaker, the F0 value was arranged according to different tones, and a tone-dependent histogram was produced.

2.4. Results and discussion

Table 2. *A confusion matrix of Mandarin tones in disyllables produced by Japanese learners.*

target	Produced as			
	Tone1	Tone2	Tone3	Tone4
Tone1	91.67%	3.12%	0.21%	5.00%
Tone2	5.74%	83.89%	6.11%	4.26%
Tone3	2.14%	52.38%	41.67%	3.81%
Tone4	4.37%	0.42%	0.42%	94.79%

Table 3. *Accuracy of Tone 3 in disyllables produced by Japanese learners. The table shows the total accuracy of all tones and the accuracy of Tone 3 in the first and second syllable.*

Speakers	JF1	JF2	JF3	JF4	JF5	JF6
Mandarin experience (Year)	1.00	1.25	2.00	2.75	3.41	4.00
All tones (%)	94.2	65.6	49.4	80.6	77.5	77.5
Pre Tone 3 (%)	80.0	46.7	6.7	13.3	13.3	26.7
Post Tone 3 (%)	80.0	0.0	13.3	20.0	6.7	6.7
Speakers	JM1	JM2	JM3	JM4	JM5	JM6
Mandarin experience (Year)	0.25	0.58	1.67	0.75	2.67	3.50
All tones (%)	72.2	74.2	92.5	90.0	89.4	90.0
Pre Tone 3 (%)	33.3	60.0	73.3	40.0	80.0	40.0
Post Tone 3 (%)	20.0	26.7	80.0	60.0	86.7	60.0

The result of Japanese learners' tone accuracy from this experiment, given in Table 2, is generally consistent with previous studies. Tone 3 is the most error-prone one and most Japanese learners tend to produce Tone 3 as Tone 2.

Table 3 shows that the overall tone accuracy increased with the speakers' Mandarin experience, but the accuracy of Tone 3 does not have this tendency. Some learners with Mandarin experience more than 2 years still cannot grasp the main feature of this tone, which reflecting the influence of fossilization phenomenon.

The more detailed error pattern can be observed from Figure 1, which displays the averaged z-scores of Tone 3's pitch contour in disyllables by the two groups. From the comparison, it should be noted that when Tone 3 appears in the first syllable, it is a "half third tone" with a low-dipping contour for native speakers; while, Japanese students' Tone 3 in this position presents as a level tone with a slight rising tendency. In the second syllable, the inflexion point and ending point of Japanese Tone 3 are much higher than the ones of Chinese, forming a contour similar to Tone 2 which is the reason of their common confusing of these two tones in disyllable words.

The overall difference between these two groups can be observed from the comparison of pitch contours. But the relative position of each tone within an individual speaker's F0 range, the consistency of native speakers and the detailed error patterns in Tone 2 and Tone 3 of Japanese learners cannot be displayed in the pitch contour figure. So, the tone-dependent F0 distribution histograms were adopted to reveal such information.
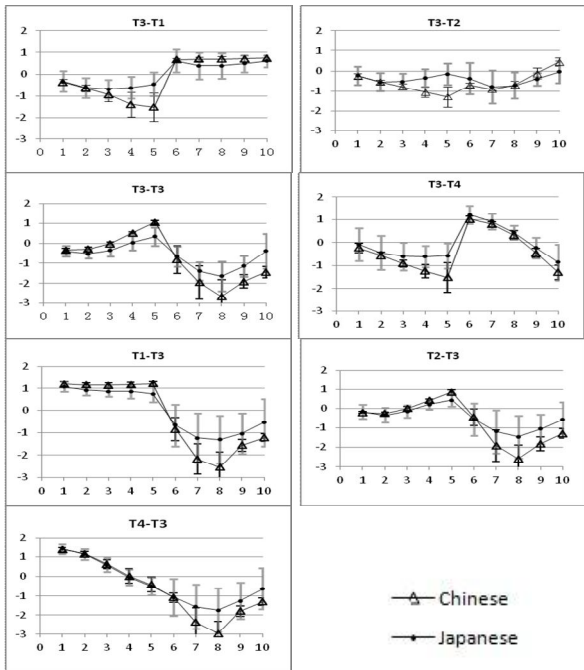
Figure1: *Chinese-Japanese pitch contour of Tone 3 in disyllables. The error bars show the standard deviation for each point. The transverse axis refers to normalized time and the unit of the vertical axis is z-score.*

Figure 2 and 3 illustrate the tone-dependent F0 distribution in disyllables of three native Chinese, which shows a systematic consistency.

- F0 distribution of native speakers has a double-peak pattern within F0 range of an individual speaker, with Tone 3 in the low and Tone 1 in the high area. The double peaks may also correspond to the high and low categories in tone perception.
- For Tone 3, native speakers' F0 has a central distribution at the low region, within 0-50% of one's tonal F0 range. This tone extends to the lowest range and does not overlap with Tone 2 in that region.
- For Tone 2 in disyllables, native speakers' F0 distribution reveals the dynamic range of the tone contour. The beginning point of F0 distribution shows a slight inter-speaker difference, but there's an obvious boundary between Tone 2 and Tone 3 in low-F0 region. As to the ending point of this tone, different speakers show a consistent tendency with F0 distribution extending to around 70% of one's F0 range.

Figure 4 and 5 demonstrate those of Japanese speakers. Although their F0 distribution in general also shows a double-peak pattern, Japanese learners with different tone accuracy have their own characteristics.

Figure 4 demonstrates the disyllable F0 distribution of a female Japanese speaker (JF3) who has a relatively low accuracy for Tone 3 with 6.7% in the first syllable and 13.3% in the second syllable, and almost all the errors are caused by the replacement of Tone 3 with Tone 2. Comparing with the native speakers, the overlap of Tone 2 and Tone 3 is obvious here, with Tone 3's F0 distribution region covering 0-72% of this speaker's F0 range, and Tone 2, 2.5%-80%. Other Japanese students with low accuracy show the same tendency (Figure 5).
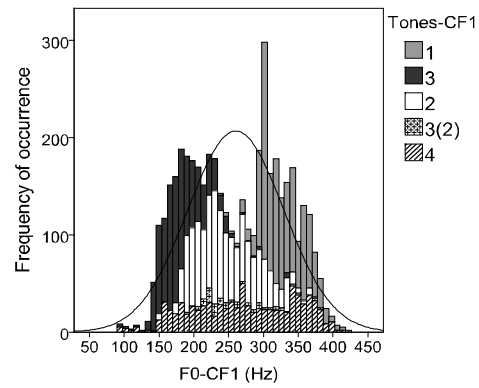

Figure 2: *F0 distribution histogram in disyllable of Chinese native speaker CF1. The number 1-4 in the histogram refers to Tone1-4 respectively and 3(2) represents the phonological changed form of Tone 3 before another Tone 3. In her range, Tone 2 covers the area of 18%-75%, and Tone 3, 0%-37%.*
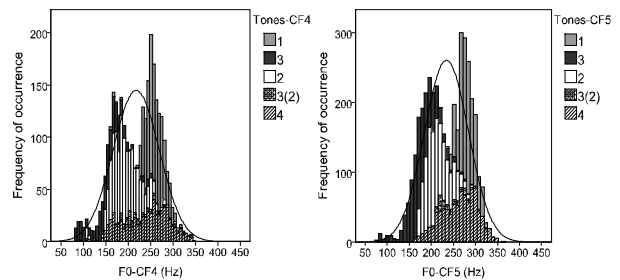

Figure 3: *F0 distribution histogram in disyllable of Chinese native speakers CF4 and CF5.*
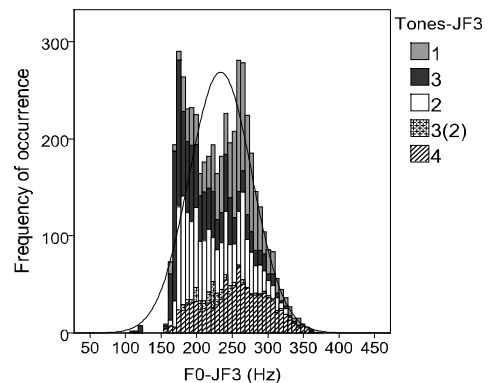

Figure 4: *F0 distribution histogram in disyllable of Japanese learner JF3 with Tone 2 covers the area of 2%-80% and Tone 3, 0%-72%.*
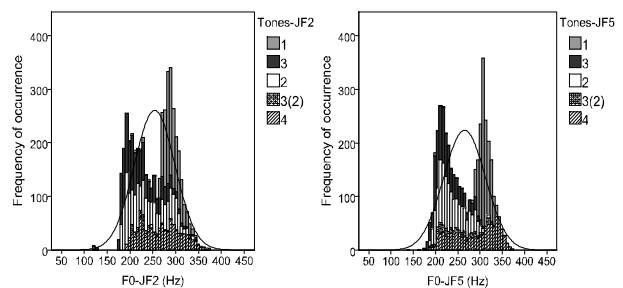

Figure 5: *F0 distribution histogram in disyllable of Japanese student (JF2, JF5) with low accuracy for Tone 3.*
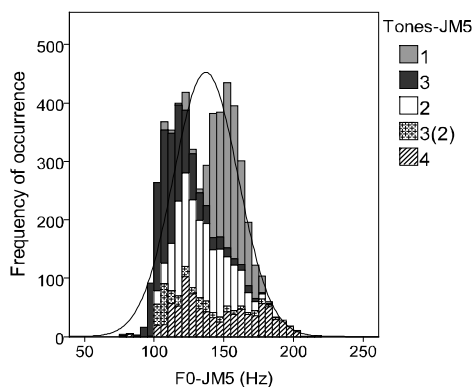
Figure 6: *F0 distribution histogram in disyllable of Japanese student （JM5） with high accuracy for Tone 3. Tone 2 covers the area of 16%-71%, and Tone 3, 0%-48%.*
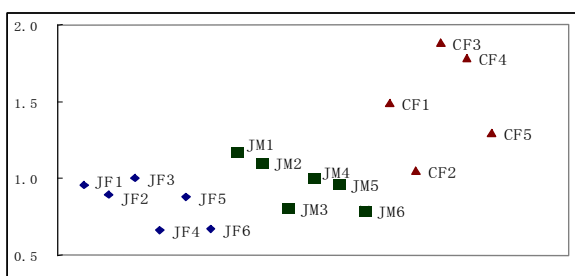

Figure 7: *Tonal range in octaves of disyllables of Chinese and Japanese learners. Here, CF and JF refer to Chinese and Japanese female, JM for Japanese male.*

Fig. 6 reveals the disyllable F0 distribution of a Japanese male learner, JM5, who has a higher accuracy (over 85% for both positions) of Tone 3. It is clear that his F0 pattern is more similar to the native speakers'. The boundary between Tone 2 and Tone 3 is clearly revealed with Tone 3 covering 0-48% of his F0 range and Tone 2, 16%-71%. This learner has acquired the "low" feature of Tone 3. Other learners with high accuracy also show the similar tendency.

Some studies demonstrated that the register range of Chinese speaker was around one octave in normal condition [6, 14]. The measurement (Figure 7) in this experiment shows that the tonal ranges of Chinese female speakers are wider than one octave, but the Japanese learners generally have a relatively narrow range when speaking Mandarin (Mean±s.d.: Chinese= 1.50±0.34; JM= 0.97±0.27; JF=0.84±0.14). The narrow tonal range of Japanese learners constrains the realization of the "low" feature in Tone 3, which is essential in the distinction of Tone 2 and Tone 3.

3. Conclusions

The main error pattern of Japanese learners in Mandarin disyllables has been clearly presented in the figures of overall pitch contours and tone-dependent F0 histograms, that is, Japanese learners tend to have a narrower tonal range and the "low" feature in Tone 3 is relatively difficult for them to acquire. In computer assisted pronunciation training system, this method can also be used to present and assess tone production of different learner groups with quantitative

information of tone contours and F0 distribution of each tone within individual speaker's F0 range.

4. Acknowledgements

5. References

[1] Norman, J, Chinese, Cambridge: Cambridge University Press, 1987.

[2] Jin, J.H., "The crucial reason and solution of problems in Chinese production by Japanese learners", Journal of Shandong Teachers' University, 5: 100-102, 1996.

[3] Tian, Y., "Teaching Japanese students Mandarin tones based on analyses of production errors", MA thesis of Tianjin Normal University, 2003.

[4] Chao, Y.R., Tone and intonation in Chinese, Bulletin of the Institute of History and Philology Academia Sinica, 4: 3, 1933.

[5] Chao, Y.R., Mandarin Primer, Harvard University Press, 1955.

[6] Lin, T, and Wang, L.J., Course of Phonetics, Beijing: Press of Beijing University, 1991.

[7] LI, C.N. and Thompson, S.A., "The acquisition of tone in mandarin-speaking children", Journal of Child Language, 4: 185-199, 1977.

[8] Yuko, Ota., "A study of production and perception of tone sandhi of Chinese disyllables by Japanese students", MA thesis of BLCU, 2011.

[9] Yang, B., "A model of Mandarin tone categories--a study of perception and production", PhD dissertation of University of Iowa, 2010.

[10] Zhang, H., "Analyses of Japanese students' Chinese tones acquisition and error", MA thesis of Fujian Normal University, 2003.

[11] Zhang, J.S. and Hirose, K., "Tone nucleus modeling for Chinese lexical tone recognition", Speech Communication, 42: 447-466, 2004.

[12] Howie, J.M., Acoustic studies of Mandarin vowels and tones, Cambridge: Cambridge University Press, 1976.

[13] Boersma, P. and Weenink, D., "PRAAT: Doing phonetics by computer," (Version 5.1.43), Online: http://www.praat.org/, accessed in 2010.

[14] Wu, Z.J, "Further experiments on spatial distribution of phrasal contours under different range registers in Chinese intonation", International Symposium on Prosody, Yokohama, Japan, 1994.